

# How to group wireless nodes together?

A survey on Matchings and Nearest Neighbour Graphs

Anastasios Giovanidis

CNRS & Télécom ParisTech - LTCI

email: anastasios.giovanidis@telecom-paristech.fr

## CONTENTS

<b>I</b>	<b>Introduction</b>	<b>2</b>
I-A	Dynamic Clusters . . . . .	2
I-B	Static Clusters . . . . .	3
I-C	Static Clusters based on Node Proximity . . . . .	3
I-D	Open Questions for Static Clustering . . . . .	4
<b>II</b>	<b>Matching Problems</b>	<b>5</b>
II-A	The Stable Marriage Problem . . . . .	5
II-B	The Stable Roommates Problem . . . . .	5
II-C	Acyclic Networks . . . . .	6
II-D	Example of Matching pairs in an Acyclic Network . . . . .	7
II-E	Matchings on infinite sets and Point Processes . . . . .	8
<b>III</b>	<b>Clusters beyond just Pairs</b>	<b>11</b>
<b>IV</b>	<b>Clusters for Point Processes</b>	<b>12</b>
<b>V</b>	<b>Clusters in the nearest-neighbour (NN) graph</b>	<b>14</b>
V-A	Properties and Results . . . . .	14
V-B	The generation number . . . . .	18
<b>VI</b>	<b>The ancestor number</b>	<b>22</b>
VI-A	Definition and Bounds . . . . .	23
VI-B	Interpretation of the ancestor number: Cluster size distribution . . . . .	30
	<b>References</b>	<b>33</b>

## Abstract

This report presents a survey on how to group together in a static way planar nodes, that may belong to a wireless network (ad hoc or cellular). The aim is to identify appropriate methods that could also be applied for Point Processes. Specifically matching pairs and algorithms are initially discussed. Next, specifically for Point Processes, the Nearest Neighbour and Lilypond models are presented. Properties and results for the two models are stated. Original bounds are given for the value of the so-called generation number, which is related to the size of the nearest neighbour cluster. Finally, a variation of the nearest neighbour grouping is proposed and an original metric is introduced, named here the ancestor number. This is used to facilitate the analysis of the distribution of cluster size. Based on this certain related bounds are derived. The report and the analysis included show clearly the difficulty of working in point processes with static clusters of size greater than two, when these are defined by proximity criteria.

## I. INTRODUCTION

In this overview report, a certain number of questions is identified that has been raised from our previous work [GKB12], [BG13], [BG15], [GACD15] related to cooperation in wireless (cellular) networks. Specifically we investigate the question of what is the best way to group nodes of a wireless network together, based on their location only. Furthermore, we propose possible steps for future research by identifying tools that exist in the literature.

Apart from the literature survey, the report contains a certain number of novel results on the Nearest Neighbour Graph and the distribution of the size of groups (from now on called clusters). These are found in Section V-B and VI.

### A. *Dynamic Clusters*

A key problem raised in the modelling and analysis of cooperative wireless networks in [BG15], [BK15], [PGH08], [NMH13], [NMH14], [TSAJ14], [LMJLH15], [SH14] has been the fact that these clusters were formed in a dynamic way. This is often not desirable, because it requires high flexibility in the inter-cell communication, as well as a large amount of information exchange. Additionally, when the same station takes part in different clusters, it should share its time or frequency resources among them, thus leading to spectral efficiency degradation. Such assumptions on dynamic cluster formation have further created problems in the analysis of interference. As a result in [BG15] the far-field approximation had to be made. Furthermore, in the same work, due to the appearance of secondary users, the same Base Station that serves the user, may also create first order interference, in order to serve the secondary users simultaneously.

Some of these observation were also identified in [AJ12], [AJ13], where the authors use a clustering method based on random centres distributed themselves as a point process and in [PLH16] where the authors provide an approach towards solution by use of graph coloring.

### B. Static Clusters

To avoid such problems for good, an idea, which we investigate here, is to model cooperative networks with disjoint and fixed clusters. In such case, a base station will only take part in at most one single cluster. Furthermore, planar areas will be permanently assigned to a specific cluster.

Analysis in this way will be facilitated and the results will give more clear evidence of possible performance benefits by clustering Base Stations and sharing their users. What is more, in such settings more complicated Multiple-Input-Multiple-Output (i.e Multiple BSs Multiple Users) cooperation scenarios can be examined, something which was not easy in the previous works, e.g. in [BG15] we analysed  $2 \times 1$  pairs of BSs that serve exactly one user/location. Other authors [AJ12], [AJ13], [LMJLH15], have made different efforts with approximative models.

### C. Static Clusters based on Node Proximity

We are interested to find grouping methods based on node proximity; hence we do not consider here channel characteristics. The reason is that we search for clusters of fixed size and fixed elements (nodes), whose choice does not depend on the variable parameters of the telecommunication systems, as for example the fading or the user positions. In this sense, our work here aims at *network-defined and fixed* clusters, to differentiate from a user-driven selection, done in our previous work [BG13], [GB13] as well as the related research by Thiele *et al.* in [TKBH12].

The criterion of relative distances for defining clusters is reasonable for the following reasons. It is related to the path-loss factor of the channel power. Clusters should in general be designed to have a stable structure for long time-intervals, so that they can provide stable link quality. The quality of a link is best described by the path-loss if we average over the fast-fading, which varies over time. The path-loss is a decreasing function of the distance from the transmitter and it deteriorates due to the path-loss exponent. Taking this into account, small relative distances will allow BSs to exchange messages more reliably (even allow the installation of fixed optical fiber communications between them, since clusters do not change). They will also allow a user to be served more efficiently, by a larger group of BSs, whose distance from the user is small, and hence their transmitted signal will on average be received in a satisfactory power level at the user's receiver. Consequently, if a group of atoms lies close enough to each other and relatively far away from other atoms, their potential cluster serving users in cooperation will be beneficial for the network performance. On the one hand, the BSs of the cluster will transmit in cooperation without causing interference to each other and at the same time they may increase their beneficial signal through network MIMO techniques. On the other hand, these BSs will be far away from other atoms, hence the inter-cluster interference that they cause will be low. Cluster choice based on path-loss has already been studied in a previous work of one of the authors [GKB12]. In the latter work however, the user

positions were assumed known and influenced the cluster choice, since the criterion was the long-term service satisfaction above a predefined threshold.

#### D. Open Questions for Static Clustering

We identify a set of open problems, that can be summarised in the following questions:

- 1) *How do we define the static clusters given a realisation of a Point Process? In other words, which atom will cooperate with which neighbour to form clusters? What is the grouping law?*
- 2) *How large should these clusters be?*
- 3) *Once the clusters are defined, how do we separate and assign the planar areas to each cluster?*

After having answered such questions, further performance analysis is possible.

This report is organised as follows:

- In Section II, some existing results from the literature on the problem of *matching pairs* are presented. These works treat finite sets of atoms and the section includes also an example of how a stable matching (pairing) could actually work. It concludes with a presentation of the problem of matching pairs for a planar Point Process in II-E. The main question is whether the existing results can be extended to the infinite case. After that, the next relevant question, is whether we can redefine the Voronoi tessellation of the plane taking as centres not anymore the individual atoms but rather the pairs, or the linear segments that connect them.
- Section III considers the important topic of advancing beyond the concept of cooperation in pairs and suggests ways of cooperation in pairs, triplets etc... taking into consideration the concepts already discussed in the previous section.
- Section IV approaches the problem of static clustering specifically for nodes produced by a realisation of a Point Process. The section investigates ways to group nodes only based on proximity. Two methods are presented from the literature, namely the Nearest Neighbour and the Lilypond model. Common properties of the two models are presented and their differences are explained.
- Section V focuses on the Nearest Neighbour graph and provides main results from the literature, on the structure and size of such clusters. A novel theorem is included, specifically, Theorem 3, where bounds for the size of a branch of the cluster from a Nearest Neighbour graph for Poisson Point Processes are derived. These are based on the so called, *generation number*, already introduced in the literature.
- Section VI introduces another metric for the size of a cluster, formed by a variation of the nearest neighbour model. The metric is named here the *ancestor number*. This number does not count only the size of one branch of the cluster, but rather aims at approximating its total size. Based on geometric bounds and for Poisson Point Processes, the ancestor number is bounded above and below in Theorem 4. The section

concludes with bounds on the tail probability of the size of a cluster produced by this variation of the nearest neighbour graph.

## II. MATCHING PROBLEMS

The literature is very rich especially in "pairing" problems (marriage, roommates) for finite sets. These relevant problems are discussed in the following subsections.

### A. The Stable Marriage Problem

In this problem first presented and investigated by Gale and Shapley (1962) [GS62], there are exactly  $n$  men and  $n$  women who wish to be paired in a stable way. Each person (woman or man) ranks those of the opposite sex in accordance to her/his preferences for a marriage partner.

Suppose we suggest a configuration of  $n$  marriages between these men and women. The configuration is *unstable* if under it there are a man and a woman, not married to each other, but who prefer each other (according to their ranking) to their actual mates. We call this a *blocking pair* for the configuration.

An important conclusion of the paper is that "*there always exists a stable set of marriages*". In other words, we can always find a configuration that is stable, for any preference list of the users. The proof also provides an algorithm to find such a stable configuration, which may not be unique. The algorithm is based on a sequence of *proposals* and *considerations*, meaning that a girl (say) may consider a boy who proposed in some round, but can reject him in a next round when a better proposal may come. In this way, the algorithm converges to a configuration without blocking pairs.

### B. The Stable Roommates Problem

This problem is an extension of the Stable Marriage Problem, and was discussed briefly as a special case, in the original Gale-Shapley paper. It was thoroughly treated later on, in the work of Irving (1985) [Irv85]. The difference with the above, is that there exists a single set of  $n$  persons who wish to be paired with each other, with the aim e.g. to choose double rooms in a dormitory. Each person, will rank the remaining  $n - 1$  in a way that lower ranking corresponds to higher preference.

The question is again, as in the stable marriage case, whether there always exists a stable configuration of roommates. This problem, which does not differentiate between men and women is of better relation to our work, where we meet a similar problem of pairing atoms in a realization of a point process.

The algorithm suggested by Irving has two phases. The first is a sequence of suggestions and considerations, where each person rejects any poorer proposals than the one that she/he already has at hand. The result of the first phase is a reduction of the preference list of the candidates. The second phase involves the identification from this list of certain cycles of users with specific preference characteristics. These cycles will determine which

persons from the reduced preference list can further be removed. The result of the second phase is either a stable matching (pairing) or the conclusion that no such matching exists for the current problem. Note that the negative conclusion can potentially come already after the first phase. It is the case when some person is left without a proposal to consider after termination of the sequence.

The basic difference between this problem and the problem of Stable Marriage is that *the former may not always obtain a stable solution*.

### C. Acyclic Networks

The authors of the reference works [LMV<sup>+</sup>06], [GLM<sup>+</sup>07] have identified a very important feature of acyclic preference lists. An acyclic list is one where there exists *no preference cycle* between  $k \geq 3$  persons  $p_1, p_2, \dots, p_k$ , where  $p_i$  prefers  $p_{i+1}$  to  $p_{i-1}$  (modulo  $k$ ).

Specifically, they proved that *an acyclic matching preference instance (i.e. list) always has a unique stable configuration*. As a consequence *any sufficiently long sequence of active initiatives leads to the unique stable configuration*.

An *initiative* here is a probing between two pairs of some possibly suboptimal configuration, which - when *active* - results in a blocking pair (i.e. breaking of the two pairs and connection of one element of the one with some other from the second pair, which is mutually preferable).

The above results easily extend to so-called *b-matchings*, where one person has a quota  $b(p) \geq 1$  on the number of mates which he can be related with.

The importance of the result lies on the fact that, most preference lists exhibit the acyclic property. Specifically, if we use a mark  $m(i, j)$  to denote the preference of  $i$  over  $j$ , we can identify three preference categories, which are always acyclic:

- *Global preferences*, where  $m(i, j) = m(j)$  and the mark is fixed for each person (e.g. bandwidth, capacity etc) and is neighbor-independent.
- *Symmetric preferences*, where  $m(i, j) = m(j, i)$  and the marks as seen as equal from both sides of the edge of the graph. Such case is found very often in real world network problems, as when the distance, or latency is used as a mark.
- *Complementary preferences*.

The key property of the acyclic networks is that *any non-trivial acyclic preference instance always has at least one loving pair*.

A loving pair  $\{p, q\}$  is a pair between persons, such that person  $p$  is ranked first by person  $q$  and vice versa. Hence, loving pairs are unbreakable. Finding them, also provides an algorithm of finding the unique stable matching which converges in a finite number of steps.

Specifically, for any acyclic preferences instance, starting from any initial configuration  $C$ , there exists a sequence of at most  $\frac{B}{2}$  initiatives leading to the stable solution, where  $B = \sum_{p \in P} b(p)$ .

The algorithm is based simply in probing (making an initiative between) different pairs at a time, to identify blocking pairs, which will be more stable. Once a loving pair is found, it is removed from the list and the process is repeated for the remaining persons.

#### D. Example of Matching pairs in an Acyclic Network

We consider in this subsection a concrete example of an acyclic network, where we apply the results of [LMV<sup>+</sup>06] and the algorithm which converges to a stable matching in pairs. Specifically, we consider an instance of eight atoms located on the two-dimensional plane as shown in Fig. 1.

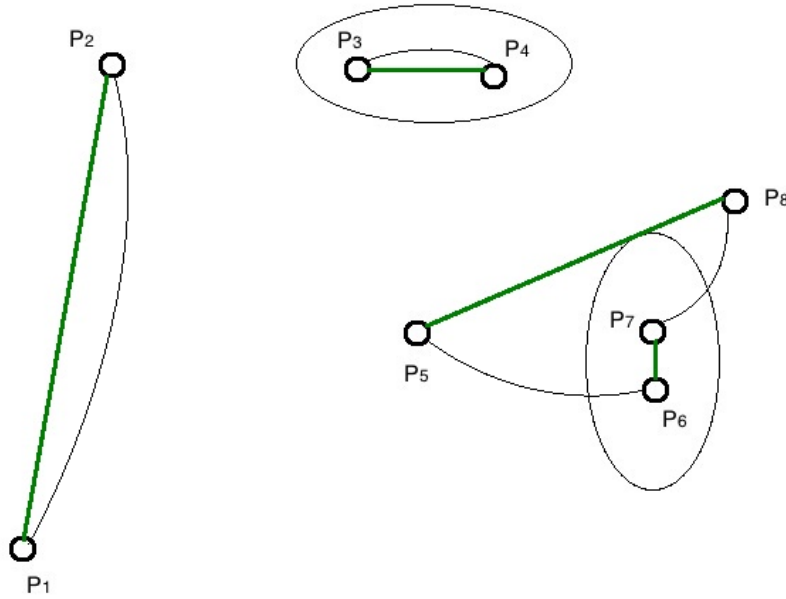


Fig. 1. Example of 8 atoms scattered on the two-dimensional plane. The black curved lines are the edges of the initial matching and the green straight lines the final optimal ones. The two pairs within the ellipses are *loving pairs*.

We can derive from the above figure the preference list of the atoms, based on their distance. Observe that when distance is used as mark, the resulting network is symmetric and consequently acyclic. Each atom ranks the rest seven atoms from 1 to 7, with 1 being its closest neighbor and 7 its farthest.

A first important observation from the preference list is that the pairs  $\{3, 4\}$  and  $\{6, 7\}$  are loving pairs, i.e. they consider each other first closest neighbours

$$P_3 \overset{1}{\leftrightarrow} P_4 \quad \& \quad P_6 \overset{1}{\leftrightarrow} P_7.$$

	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$	$P_6$	$P_7$	$P_8$
$P_1$	X	4	7	7	7	7	7	7
$P_2$	2	X	2	5	6	6	6	6
$P_3$	3	1	X	1	5	5	5	5
$P_4$	4	2	1	X	3	4	4	3
$P_5$	1	3	3	2	X	3	3	4
$P_6$	5	6	6	6	2	X	1	2
$P_7$	6	5	5	3	1	1	X	1
$P_8$	7	7	4	4	4	2	2	X

Fig. 2. Preference list for the 8 atoms of the above example, based on the distance from each other.

We initialize a constellation  $\mathcal{C}_0$  by pairing at random. The initial pairing is shown on Fig. 1 with the black curved lines between atoms. The ellipses surrounding the two pairs highlight the loving pairs in our example. We can write down the initial constellation  $\mathcal{C}_0$

$$\begin{aligned}
P_1 &\leftrightarrow P_2 && \left( P_1 \xrightarrow{2} P_2, \quad P_2 \xrightarrow{4} P_1 \right) \\
P_3 &\leftrightarrow P_4 && \left( P_3 \xrightarrow{1} P_4, \quad P_4 \xrightarrow{1} P_3 \right) \\
P_7 &\leftrightarrow P_8 && \left( P_7 \xrightarrow{2} P_8, \quad P_8 \xrightarrow{1} P_7 \right) \\
P_5 &\leftrightarrow P_6 && \left( P_5 \xrightarrow{2} P_6, \quad P_6 \xrightarrow{3} P_5 \right).
\end{aligned}$$

The notation  $P_i \xrightarrow{n} P_j$  means that the atom  $P_j$  is ranked  $n$ -th by  $P_i$ . The steps of the algorithm are as follows:

- 1) The pair  $\{P_3, P_4\}$  is a loving pair and hence can be removed from the list.
- 2) From the two pairs  $\{P_7, P_8\}$  and  $\{P_5, P_6\}$  a blocking pair can be found  $\{P_6, P_7\}$ , which is also a loving pair. Hence the above two pairs break, the loving pair  $\{P_6, P_7\}$  is formed and removed from the preference list. Furthermore, the pair  $\{P_5, P_8\}$  is formed.
- 3) The remaining pairs are  $\{P_1, P_2\}$  and  $\{P_5, P_8\}$ .
- 4) Observing the list of preferences, we see that no blocking pair can be created from the two, hence the two remaining pairs are stable.
- 5) The final matching is shown in Fig. 1 with green straight lines.

#### E. Matchings on infinite sets and Point Processes

##### 1) Infinite Sets and Matching - Basic Questions



The great *challenge* is to find out whether the results of existence of a stable matching for finite sets can be extended to the infinite case, where the atoms of a Point Process form preference lists - possibly having as criterion the distance of each one with all its neighbours. Important questions to be answered here are the following:

- Does such a stable matching *always exist* for  $n \rightarrow \infty$  and for each type of Point Process (Poisson, Mattern, etc.)?)
- Is there a *unique matching*?
- Is there a *polynomial-time* algorithm which is guaranteed to converge to a possible stable matching? Can we propose a *decentralized* solution?

In the results of the previous paragraph, we saw that for finite acyclic networks, such as those which use the distance as mark (i.e. relevant to the Point Processes and the problems of interest), there exists a *unique stable configuration*. Hence, the conjecture that a unique stable matching of pairs for  $n \rightarrow \infty$  exists, when the criterion is the distance, seems very logical (still needs to be proved).

Problems arise, only when there exists a preference cycle between  $k \geq 3$  atoms  $p_1 \dots p_k$ , where  $p_i$  prefers  $p_{i+1}$  to  $p_{i-1}$ . But this is impossible due to the distance ranking.

The special case, where  $k \rightarrow \infty$  should not be excluded. This is the case where each atom  $p_i$  is the closest neighbour for  $p_{i-1}$ , but its own closest neighbour  $p_{i+1} \neq p_{i-1}$  and lies outside the ball with center  $p_{i-1}$  and having  $p_i$  on its boundary (repeated infinitely many times). Such cases of atom placement are relevant with the so called *lilypond model*, first appearing in [HM96] for the Poisson Point Process and further studied by Last and Penrose (2010) [LP10].

Specifically, it is proven in [LP10, Lemmas 2.1, 2.2] under certain assumptions, that no such infinite preferential cycle of atoms can exist.

## 2) Further Questions related to Geometry

Relevant further questions - after the above have been answered - are related to the *distribution of geometric characteristics* of the matchings on the two-dimensional plane.

- What can we say about the distribution of the distance between two matched partners?
- Can we define a new *tessellation of the plane* - like the 1-Voronoi tessellation - which is related to the line segments of matched pairs?
- How does the distribution of neighbours around an atom affect the results on matching? (mutual neighbour distribution)

## 3) Voronoi Tessellation of Segments

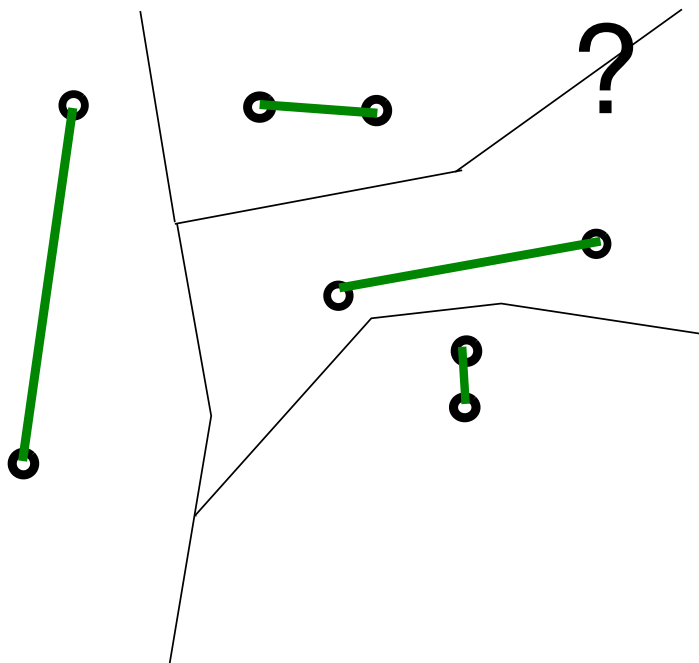


Fig. 3. Example of how the tessellation of the 2-dimensional plane could look like, taking as centers the linear intervals defined by the previous matching. We show again the example of the eight atoms similar to the previous section.

Especially the second question in the previous subsection is of particular interest. After determining the appropriate matchings, we would like to separate the plane into *compact* and maybe *convex* subregions, each one of which should be related to exactly one cooperation cluster. An example is shown in Fig. 3.

By connecting the atoms of each cluster, line segments and polygons are formed, and the question on the appropriate tessellation is now stated more generally.

As shown in [FEC02], the Voronoi cell definition is not restricted to distinct points, when referring to sites. So, the concept of the Voronoi cell can be generalized to sets of points like e.g. the matched atom pairs, or the linear intervals. Furthermore, an interesting concept relevant to the Voronoi tessellation for segments, is the so-called *medial axis*, and the relation between these two notions.

Another relevant publication, which extends (or rather provides an alternative notion to) the Voronoi tessellation, is the work by Hoffman, Holroyd and Peres (2006), which relates planar points to cells of atoms in a more "fair" way, so that all subregions have equal volume [HHP06]. The volumes are defined by a so-called appetite, modeled by the parameter  $\alpha \in (0, \infty]$  and the points are assigned to atoms (and relevant Voronoi cells) based on the Gale-Shapley stable marriage principle. When  $\alpha \rightarrow \infty$  the tessellation is identical to the Voronoi tessellation of the plane.

### III. CLUSTERS BEYOND JUST PAIRS

It is reasonable to continue our research beyond the constraint that only pairs of Base Stations are allowed as clusters and all atoms should belong to exactly one such pair. This is a *very restrictive assumption that does not necessarily model the real systems and how cooperative clusters might work within them.*

A different approach would be to allow pairs, triplets, quadruplets, etc. of atoms cooperate, while at the same time certain atoms might not cooperate at all. This approach sounds more reasonable, when atoms form cooperative entities with increasing size depending on their distance and only when this is necessary and helpful. In other words, it might not pay much to form a cooperative pair between two atoms that lie very far away from each other, just for the sake of forming a pair.

A possible way to construct such clusters will be as follows:

- 1) Start with the realization of a Point Process. The current constellation is the *non-cooperative case*.
- 2) Identify pairs of atoms that are *mutual first neighbours* to one another as in [GACD15].

$$P_i \overset{1}{\leftrightarrow} P_j.$$

These neighbours will form the pair-clusters and the new constellation will be called *2-cooperative case*. Of course all the atoms that do not find a pair with mutual first neighbour will be treated as single atoms, so that pairs and single Base Stations coexist.

- 3) Among the cooperative pairs, identify triplets of atoms, such that the following chain of preference (related to distance) holds

$$P_{i-1} \overset{1}{\rightarrow} P_i \overset{1}{\leftrightarrow} P_{i+1} \quad \text{or} \quad P_{i-1} \overset{1}{\leftrightarrow} P_i \overset{1}{\leftarrow} P_{i+1}.$$

These obviously constitute the *3-cooperative case*. In such scenario single atoms, pairs and triplets coexist. In the formation of such triplets, possibly a constraint on the distance should be considered, so that cases where a third single atom which is far away is included in the triplet are avoided.

(Notice: Remember that the network is acyclic so we do not consider the third case where  $P_{i+1}$  may have first preference to  $P_{i-1}$  or the other way round. Also, the case of three equidistant atoms is impossible and has zero probability in the Poisson Point Process case.)

- 4) Among the cooperative triplets identify chains of preference with four atoms, and so on...
- 5) ...

This type of clustering is based and builds on the previous definition of *loving pairs* in the subsection for acyclic networks. The clusters are formed as cooperative entities including atoms which have only relations of first preference (first closest neighbour) at least in one direction. An example of such a clustering is shown in the following Fig. 4.

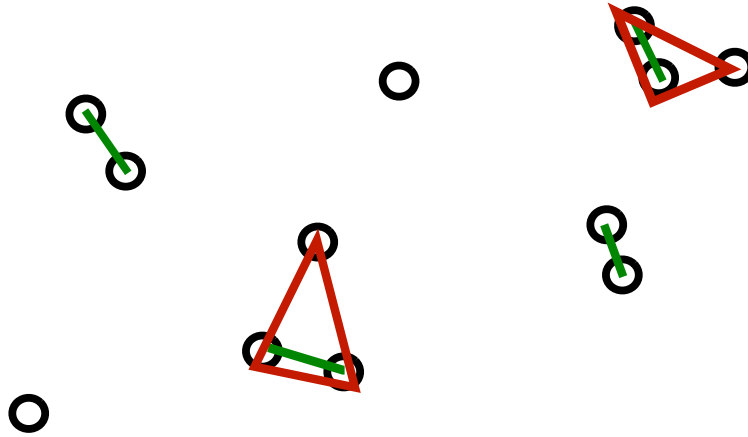


Fig. 4. Example of the proposed matching with no-cooperation, 2-cooperation and 3-cooperation case. The pairs are shown with green linear intervals and the triplets with red.

After such clusters are formed consideration of Voronoi tessellations follow. Finally each case of 2-, 3-,... cooperation is compared to the case of no cooperation in order to derive possible benefits related to performance measures such as coverage and throughput.

Other types of cluster formation are also under discussion.

The direct benefit of such an approach is however that the clusters are isolated entities and each one of these has a defined planar area of interest/association. There is no overlap either of planar Voronoi cells or of atoms that might belong to the same cluster. This helps a lot in the analysis and models a way of function that has been suggested as appropriate for cooperative communications systems.

#### IV. CLUSTERS FOR POINT PROCESSES

Let us consider a homogeneous point process (p.p)  $\Phi$  in  $\mathbb{R}^2$  with non-negative density  $\lambda > 0$ . When referring to a Poisson Point Process we will use the abbreviation p.p.p. One realization of the process is  $\phi$  and can be described by the infinite set of atoms  $\{\mathbf{z}_i\}$ , where  $\mathbf{z}_i = (x_i, y_i)$ . Each realization shows a possible deployment of single antenna Base Stations (BSs) on the plane. We wish to group these BSs (or atoms) into *disjoint cooperative clusters*, with possibly different sizes, where size here means the cardinality of atoms included in each cluster. To be more formal, a cluster is defined to be a finite subset  $\mathcal{C}(\phi)$  of the realization  $\phi$  (we omit  $\phi$  from now on unless we refer to different realizations of the p.p.), with cardinality  $\text{card}(\mathcal{C})$ . Exactly as the atoms of  $\phi$  are somehow enumerated and indexed by  $i$ , we enumerate and index the clusters using the index  $m$ . We consider clusters of atoms with the following two properties: (a) the set of all clusters constitutes a partition of  $\phi$ , hence

their union exhausts the infinite set of atoms and (b) they are disjoint subsets of  $\phi$ , meaning that the intersection between any pair of distinct clusters is empty

$$\bigcup_{m=1}^{\infty} \mathcal{C}_m = \phi, \quad (1)$$

$$\mathcal{C}_m \cap \mathcal{C}_n = \emptyset, \quad \forall m \neq n. \quad (2)$$

For the partition of  $\phi$  into clusters, we intend to use rules that depend only on the geometry. In other words, given a realization of the p.p., an atom  $\mathbf{z}_i$  will take part in a cluster, based only on its relative distance to the rest of the atoms  $\phi \setminus \{\mathbf{z}_i\}$  as well as their own relative position.

We consider in this report different ways to connect the atoms of a p.p. and define the cooperative clusters, based solely on  $\phi$ . A connection between two atoms means that these cooperate and belong to the same cluster. These two models have been presented and studied partly in the work of Häggström and Meester [HM96]. Specifically we will use:

- 1) the **nearest neighbour (NN) model**. Given the realization  $\phi$  we connect each  $\mathbf{z}_i$  to its *nearest neighbour* by an undirected edge. This results in a graph  $\mathcal{G}_{NN}$ , which is well defined at least for a p.p.p. where no two inter-atom distances are the same a.s. and hence each atom has a unique first neighbour. However, an atom can be a nearest neighbour for a set of atoms (possibly empty).
- 2) the **lilypond (LL) model**. We construct the graph  $\mathcal{G}_{LL}$  dynamically as follows. Starting by the realization  $\phi$ , we assume at time  $t = 0$  that there is a ball of radius  $t$  (hence 0) centered on each atom  $\mathbf{z}_i$ . Then we let time  $t \uparrow$  evolve and the radii of these balls grow (linearly in  $t$  and all with the same speed). As soon as a ball hits another ball, it stops growing forever. Notice that the other ball could be itself in a phase of growth (or not). The time instant that the hitting takes place, say  $t_n = t$ , is saved and gives the ball's radius  $r_i := t_n$ . In this way, a sequence of ball-touching times (equivalently of radii) is formed  $t_1, t_2, \dots, t_n, \dots$  and we are interested in the limiting configuration as  $t \rightarrow \infty$ . Each time instant a ball touches another one and stops growing, an undirected link is drawn at the graph  $\mathcal{G}_{LL}$ . The lilypond model falls in the category of the hard sphere Boolean model, where balls of different radii centered at atoms of the p.p.p. do not overlap but are possibly tangential.

In general the two models result in different partitions of  $\phi$  and different sets of clusters. There are however certain interesting *properties* (**P.x**) shared by both models:

- I) The cluster formation is independent of the density  $\lambda$  of the p.p.p..
- II) The graph  $\mathcal{G}$  (either  $\mathcal{G}_{NN}$  or  $\mathcal{G}_{LL}$ ) is disconnected, i.e. there exist two atoms that are not connected by any path.
- III) Each resulting cluster  $\mathcal{C}$  does not contain *cycles*, it is a *tree* and hence the graph  $\mathcal{G}$  is a *forest*.

- IV) The graph  $\mathcal{G}$  contains a.s. no infinite component, i.e. it does not percolate [HM96, Th.2.1 and Th.5.2]. Consequently, the cardinality of each cluster  $\text{card}(\mathcal{C})$  is a.s. finite.
- V) There exists no isolated node in the graph  $\mathcal{G}$ , i.e. there exists no cluster of  $\text{card}(\mathcal{C}) = 1$ . For the case  $\mathcal{G}_{NN}$  this is because all atoms have a nearest neighbour. For the case  $\mathcal{G}_{LL}$  because each ball will eventually touch another one as  $t \rightarrow \infty$  (the probability of an empty ball for p.p.p. is  $e^{-\lambda\pi t^2} \xrightarrow{t \uparrow} 0$ ).

We first show in Fig.5(a)-(b), the  $\mathcal{G}_{NN}$  and  $\mathcal{G}_{LL}$  graphs produced when applying the nearest-neighbour and lilypond rules on an example realization of a p.p.p. with density  $\lambda = 2$  [atoms/ $m^2$ ]. Furthermore, we better explain how the  $\mathcal{G}_{LL}$  graph is formed as the balls of radius  $t$  increase and touch each other. For this, we illustrate four instants of their growth (including the end instant) in Fig.6(a)-(d).

Although the two graphs share many common properties as presented above, it is obvious from the subfigures Fig.5(a)-(b) that they actually are different. Of course they have similarities, however there clearly are cases where different clusters are formed. To give an example, let us focus on the largest cluster at the right handside of the plane denoted here by  $\mathcal{C}_o$ , which in the NN graph has cardinality  $\text{card}(\mathcal{C}_o^{(NN)}) = 7$  and in the graph LL has cardinality  $\text{card}(\mathcal{C}_o^{(LL)}) = 5$ . A doublet in the lilypond model appears as an independent cluster of cardinality 2, although in the NN graph it belongs to the large cluster. This occurs because in the LL model an edge of the graph may connect atoms that are not necessarily nearest neighbours. This is illustrated for clarity in Fig.7. For the atom  $(2, 0)$  its nearest neighbour is actually  $(2, 1.5)$  (inside the dashed circles). However, the latter is linked with its own nearest neighbour  $(2.6, 1.5)$ , while due to the growth process of the LL model, the balls centered at  $(2, 0)$  and  $(2, 1.5)$  do not touch each other, due to the presence of the atom at the origin  $(0, 0)$ .

In the paragraphs below, we will analyze the two models separately. Furthermore, we will provide results related to their size and probability, together with a comparison between them.

## V. CLUSTERS IN THE NEAREST-NEIGHBOUR (NN) GRAPH

### A. Properties and Results

As discussed previously, all clusters  $\mathcal{C}_m$  in the NN model are a.s. finite. This was presented in property **P.IV** and holds due to the fact that there exists no infinite component. We randomly choose an atom and set its position as the origin of the planar coordinates. This atom is called the *typical atom* and is denoted by  $\mathbf{z}_o$ . Let us further denote by  $\mathcal{C}_o$  the *typical cluster*, or else the cluster in which the typical atom belongs to. Then, starting from the typical point we form the sequence of atoms  $\mathbf{z}_o \xrightarrow{1} \mathbf{z}_1 \xrightarrow{1} \dots \mathbf{z}_n \xrightarrow{1} \mathbf{z}_{n+1} \xrightarrow{1} \dots$ , where the  $(n+1)$ -th atom is the nearest neighbour of the  $n$ -th one. Hence  $\{\mathbf{z}_o, \mathbf{z}_1, \dots\} \subseteq \mathcal{C}_o$ . From the distances  $r_n := |\mathbf{z}_n - \mathbf{z}_{n+1}|$ , we obtain the sequence  $\{r_o, r_1, \dots\}$ . To facilitate the analysis, we introduce the *directed NN graph*  $\vec{\mathcal{G}}_{NN}$  where  $\mathbf{z}_i, \mathbf{z}_j$  are connected by an arc if  $j$  is the nearest neighbour of  $i$ . By replacing the arcs by undirected edges we get  $\mathcal{G}_{NN}$ . The introduced sequence of atoms makes more sense in the directed graph.

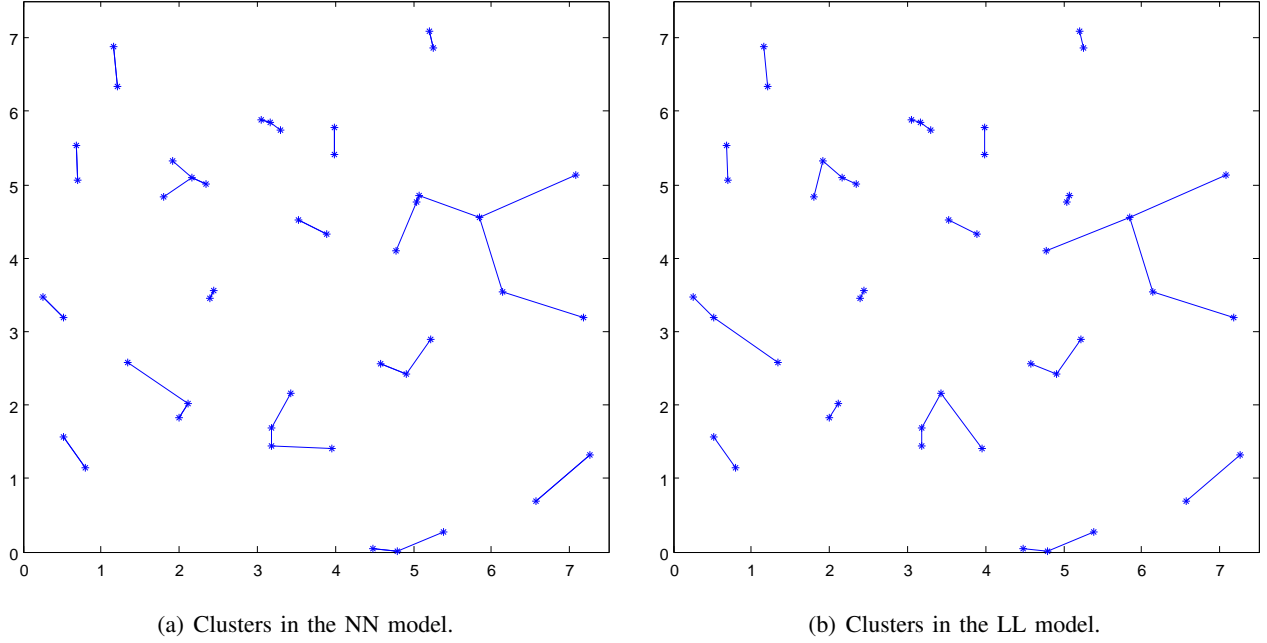


Fig. 5. Example of cooperation clusters for the same example node topology. (a) Nearest-neighbour (NN) (b) Lilypond (LL).

In the NN model, a very important subset of atoms of  $\phi$  are the *mutually-nearest-neighbours*, which are pairs of atoms with the property that the one is the nearest neighbour of the other and vice-versa. Their importance is better understood by the following two Lemmas.

**Lemma 1.** *All clusters of the NN model contain exactly one pair of mutually-nearest-neighbours.*

*Proof.* Suppose there is no pair of mutually-nearest-neighbours in a cluster, say the typical one  $\mathcal{C}_o$ . Then, the sequence of atoms  $\{\mathbf{z}_o, \mathbf{z}_1, \dots\}$  described above will either be infinite (impossible by property **P.IV**) or will be finite with  $N < \infty$ . In the second case, the nearest neighbour of  $\mathbf{z}_N$  is some atom previously encountered in the sequence with index  $n \leq N - 2$  and a cycle will appear in the cluster  $\mathcal{C}_o$  of the graph  $\vec{\mathcal{G}}_{NN}$  (and  $\mathcal{G}_{NN}$ ), with size greater or equal to 3. This is impossible (see also **P.III**).

To see why, choose without loss of generality (w.l.o.g.)  $\mathbf{z}_N \xrightarrow{1} \mathbf{z}_{N-2}$ . Then the open ball  $\mathcal{B}(\mathbf{z}_N, r_N)$  with center  $\mathbf{z}_N$  and radius  $r_N := |\mathbf{z}_N - \mathbf{z}_{N-2}|$  will be empty with  $\mathbf{z}_{N-2}$  on its boundary. Around  $\mathbf{z}_{N-2}$  there should be an empty open ball of radius  $r_{N-2}$  with  $\mathbf{z}_{N-1}$  on its boundary, and for this to hold  $r_{N-2} < r_N$ . Given that  $\mathcal{B}(\mathbf{z}_N, r_N)$  is empty, the atom  $\mathbf{z}_{N-1}$  should lie outside this ball, hence at a distance  $r_{N-1} > r_N$ . However, this contradicts the fact that  $\mathbf{z}_N$  is the nearest neighbour of  $\mathbf{z}_{N-1}$ , because the ball  $\mathcal{B}(\mathbf{z}_{N-1}, r_{N-1})$  is not empty but contains  $\mathbf{z}_{N-2}$ , since  $r_{N-1} > r_N$  and  $r_{N-2} < r_N$ .

We have proven existence of at least one mutually-nearest-neighbours pair within each cluster. The uniqueness comes from the fact that there exists no path of atoms in  $\vec{\mathcal{G}}_{NN}$ , that connects two distinct such pairs. Atoms of

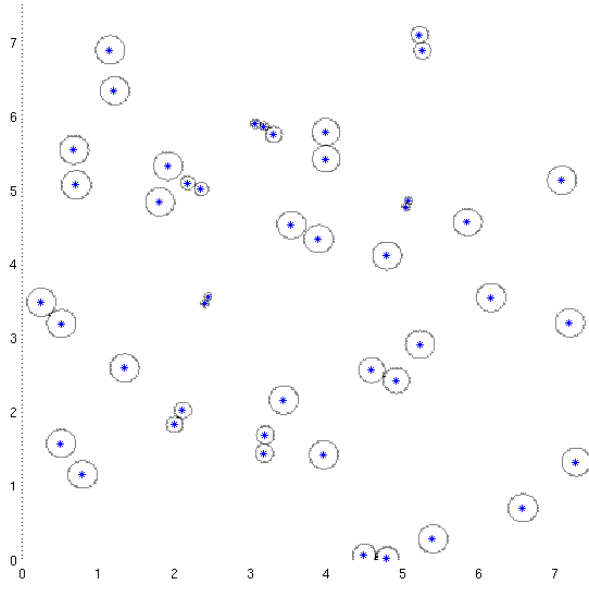
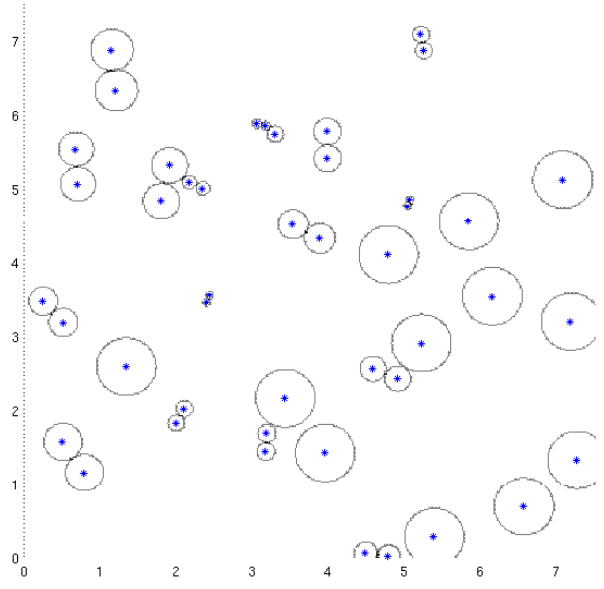
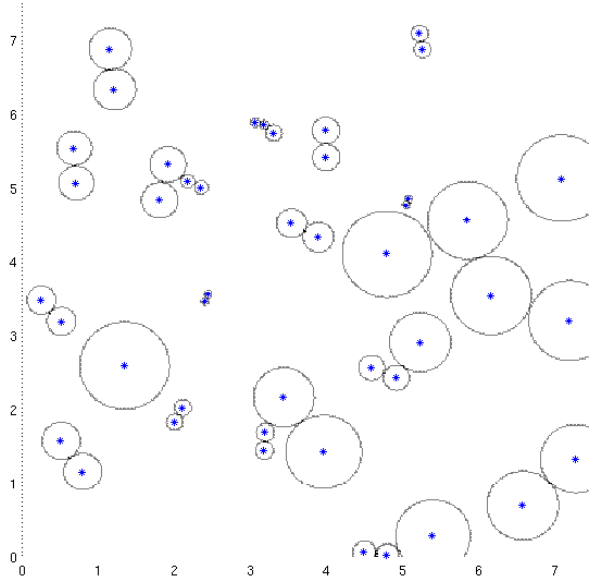
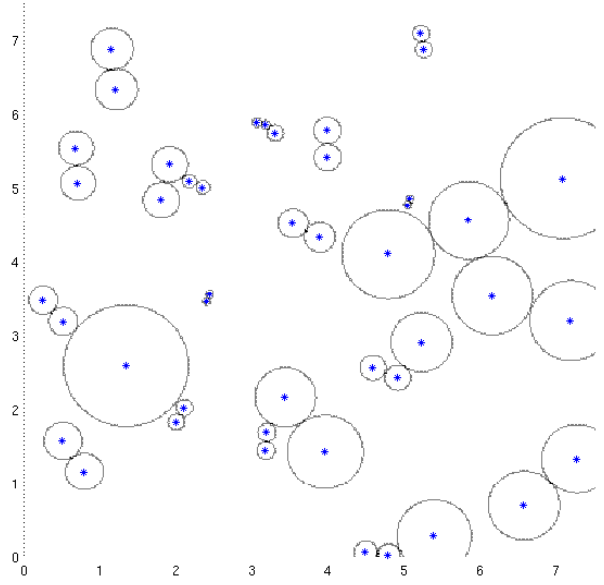
(a) Evolution of the Lilypond  $t = 0.20$ .(b) Evolution of the Lilypond  $t = 0.40$ .(c) Evolution of the Lilypond  $t = 0.60$ .(d) Evolution of the Lilypond  $t = 0.83$ .

Fig. 6. Four instants of growth for the Lilypond model using the same example topology as in previous Figure. (a)  $T = 0.20$ , (b)  $T = 0.40$ , (c)  $T = 0.60$ , (d) end of growth  $T = 0.83$ .

such pairs can only be nearest neighbours for some other atom.  $\square$

As a conclusion of the above Lemma, starting from a typical point and moving through the directed path of the graph  $\vec{\mathcal{G}}_{NN}$ , we get a sequence of  $N + 1$  points  $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_N\}$ , with  $N \geq 2$ , from which the two last ones



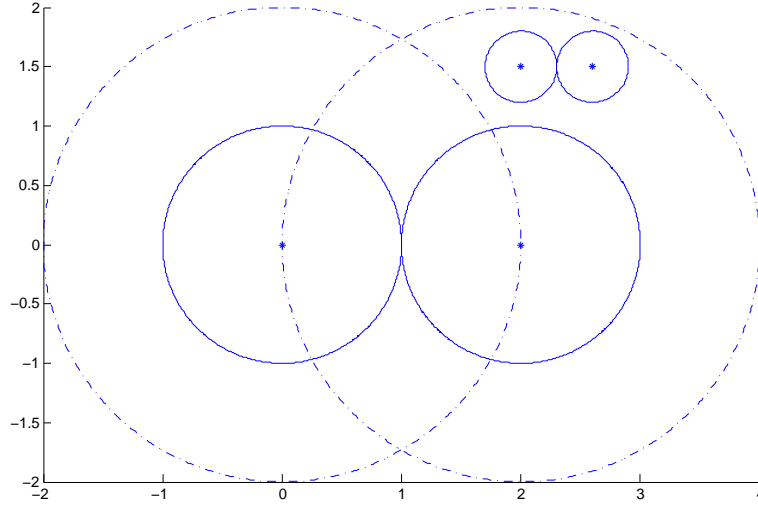


Fig. 7. Illustrative explanation of the differences between the NN and the LL graph.

$\mathbf{z}_{N-1}, \mathbf{z}_N$  are mutually-nearest-neighbours. This is the only pair of atoms of the cluster  $\mathcal{C}_o$ , having such property.

**Lemma 2.** *The distances  $\{r_o, r_1, \dots, r_N\}$  - starting from the typical atom  $\mathbf{z}_o$  - form a strictly decreasing finite sequence until the  $(N-1)$ -th atom. The sequence ends with  $r_{N-1} = r_N$ . Hence, the minimum distance is the one between the mutually-nearest-neighbours. The atom  $\mathbf{z}_N$  is called the "root" of the branch of the cluster, where  $\mathbf{z}_o$  belongs to.*

*Proof.* Starting from  $\mathbf{z}_o$ , the ball  $\mathcal{B}(\mathbf{z}_o, r_o)$  is empty, with  $\mathbf{z}_1$  on its boundary. Since these are not mutually-nearest-neighbours, the ball  $\mathcal{B}(\mathbf{z}_1, r_1)$  is empty if  $r_1 < r_o$  and has  $\mathbf{z}_2$  on its boundary, which should additionally lie outside the ball  $\mathcal{B}(\mathbf{z}_o, r_o)$ . We iterate this process until the ball of point  $\mathbf{z}_{N-1}$  with  $\mathbf{z}_N$  on its boundary. The sequence stops when  $r_N = r_{N-1}$  and the nearest neighbour of  $\mathbf{z}_N \xrightarrow{1} \mathbf{z}_{N-1}$ . An example of such a sequence of decreasing empty balls is shown in Fig.8.  $\square$

The event that the *size of a cluster* is very large may have small probability, but not zero. Note here that  $\text{card}(\mathcal{C}_o) \geq N+1 \geq 2$  in the NN graph. Regarding the asymptotic behaviour of the cluster size, Kozakova *et al.* in [KMN06] have derived bounds on its tail probability. In the case of the 2D plane of interest, we present the following result from their work.

**Theorem 1** (Theorem 1.2, [KMN06]). *Define  $\rho(N)$  as the probability that there is a directed path in  $\vec{\mathcal{G}}_{NN}$  through the typical atom at the origin, touching more than  $N$  distinct atoms. Then, there exist constants  $C_1, C_2, N_o \in$*

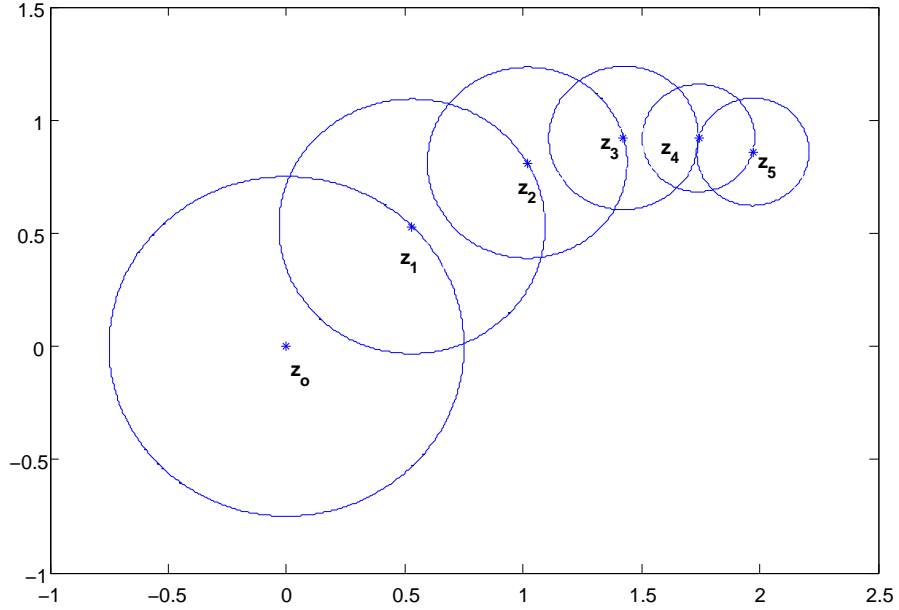


Fig. 8. Illustrative presentation of a sequence of atoms  $\{z_0, \dots, z_5\}$  in the  $\mathcal{G}_{NN}$  graph that have first neighbour relation. The figure shows the empty balls centered on each atom and the decreasing sequence of radii until the mutually-nearest-neighbour pair  $z_{N-1} \overset{1}{\leftrightarrow} z_N$ .

$(0, \infty)$  such that

$$e^{-C_1 N \log N} \leq \rho(N) \leq e^{-C_2 N \log N}, \quad \forall N \geq N_o. \quad (3)$$

Note that the elements of such path through  $z_o$  are a subset of  $\mathcal{C}_o$ . Their cardinality can be understood as the depth of the tree  $\mathcal{C}_o$ , since (as mentioned in **P.III**)  $\mathcal{G}_{NN}$  is a forest. Based on the above theorem, the probability that the length of such path inside the typical cluster  $\mathcal{C}_o$  of  $\vec{\mathcal{G}}_{NN}$  exceeds  $N$ , decays superexponentially with the number of atoms  $N$  in the path. Taking also into account the fact that the number of atoms with the same nearest neighbour is small, we conclude that the probability of very large clusters is trivial.

### B. The generation number

Another argument [KMN06] in favor of the small size of clusters in the NN model, comes from the use of the so called **generation number**. As already mentioned, any cluster of  $\mathcal{G}_{NN}$  contains exactly one pair of mutually-nearest-neighbours. These atoms are given generation number 1. A Poisson atom  $z_n$  receives generation number  $k \geq 2$ , if the graph distance (number of hops) to the unique pair in its cluster is equal to  $k - 1$ . The generation number is connected to the length  $N$  of the chain of atoms  $\{z_o, z_1, \dots, z_N\}$  with the relation  $k = N$ . To see this, if  $z_{N-1}, z_N$  are mutually-nearest-neighbours and take the generation number 1, then  $z_{N-2}$  will take generation number 2, so that  $z_{N-k}$  will take generation number  $k$ . Obviously for  $z_o$ ,  $N - k = 0 \Rightarrow N = k$ .

Let us denote by  $g(k)$  the probability that the typical atom has generation number  $k$ . We can calculate exactly the probability of generation number 1, which corresponds to the probability that a random atom of the p.p.p. belongs to a mutually-nearest-neighbour pair.

**Theorem 2.** *The exact probability that the typical atom of a p.p.p.  $\Phi$  with density  $\lambda > 0$  has a mutually-nearest neighbour (or that the typical atom has generation number 1) is equal to*

$$\mathbb{P}[\mathbf{z}_o \text{ in pair}] = g(1) = 0.6215. \quad (4)$$

*Proof.* The result has been first encountered in [DSS99] and we add its proof here for completeness. Given that the volume of the ball  $\mathcal{B}(\mathbf{z}_o, r)$  is  $V_o (= \lambda \pi r^2)$ , we can write the exact above probability as follows

$$g(1) = \int_0^\infty e^{-V_o} e^{-(1-\alpha)V_o} dV_o, \quad (5)$$

where  $(1-\alpha)V_o$  is the area  $\mathcal{B}(\mathbf{z}_1, r) \setminus \mathcal{B}(\mathbf{z}_o, r)$  outside the ball of  $\mathbf{z}_o$ . (Notice that this area is upper and lower bounded by  $V_o$  and  $V_o/2$  respectively to give the bounds in the Theorem 3 below). Furthermore,  $\alpha V_o$  is the area of overlap for the two balls of equal volume with centers at distance  $r$ . The constant is equal to  $\alpha = \frac{2}{3} - \frac{\sqrt{3}}{2\pi}$ . Having said this, the integral can be easily solved to give  $g(1) = \frac{1}{2-\alpha} = 0.6215$ .  $\square$

The probability  $g(k)$ ,  $k \geq 1$ , is upper- and lower-bounded in the following Theorem. The proof of the lower bound comes from [KMN06, Theorem 1.3], whereas the upper bound is a novel contribution.

**Theorem 3.** *Given a planar p.p.p.  $\Phi$ , the probability that the generation number, i.e. the distance of the typical atom from the mutually-nearest-neighbours, is  $k \in \{1, 2, 3, \dots\}$  can be bounded as follows*

$$2^k \frac{k+1}{(k+2)!} \geq g(k) \geq \frac{k}{(k+1)!} - \mathcal{Q}(k), \quad k \geq 1. \quad (6)$$

The correction term  $\mathcal{Q}(k) := q(k) + \sum_{i=1}^{k-1} \mathcal{Q}(i) - \mathcal{Q}(i-1)$ ,  $\mathcal{Q}(0) = 0$ ,  $\mathcal{Q}(k) \geq 0$ , is increasing in  $k$ . It holds  $\mathcal{Q}(1) = q(1) = 0$  and  $\mathcal{Q}(k) \geq \mathcal{Q}(2) = \frac{\alpha}{2}$ , where  $\alpha$  is defined as previously. The lower bound expression is equal to zero for  $k \geq 3$ .

*Proof.* Consider a typical point  $\mathbf{z}_o$  and suppose its generation number is  $k$ . Then we have a sequence  $\{\mathbf{z}_o, \mathbf{z}_1, \dots, \mathbf{z}_N\}$  of  $N+1$  atoms, among which the last two constitute mutually-nearest-neighbours. From Lemma 2 there is a sequence of balls (disks in 2D) with centers the above atoms and decreasing radii, as shown in Fig.8 for six atoms. Suppose that the decreasing sequence of their volumes is  $V_o \geq \dots \geq V_{N-1} = V_N \geq 0$ . The probability of such an event is:

- lower-bounded by the probability that the area  $\mathcal{A}^L$ , equal to the *sum of all  $N+1$  disks without overlapping*, is empty of any other atoms, reduced by a correction term  $\mathcal{Q}(N)$ , which is increasing over  $N$ . We calculate this by

$$\int_{V_o \geq \dots \geq V_{N-2} \geq V_{N-1} \geq 0} e^{-V_o} e^{-V_1} \dots e^{-V_{N-2}} e^{-2V_{N-1}} dV_{N-1} dV_{N-2} \dots dV_1 dV_o - \mathcal{Q}(N). \quad (7)$$

- upper-bounded by the probability that the area  $\mathcal{A}^U$ , equal to the *sum of the first disk  $V_o$  and all the rest  $N$  semidisks without overlapping*, is empty of any other atoms and we calculate this by

$$\int_{V_o \geq \dots \geq V_{N-2} \geq V_{N-1} \geq 0} e^{-V_o} e^{-V_1/2} \dots e^{-V_{N-2}/2} e^{-V_{N-1}} dV_{N-1} dV_{N-2} \dots dV_1 dV_o. \quad (8)$$

The lower bound is obvious, since the area of all disks without overlapping is larger than the actual area  $\mathcal{A} \subset \mathcal{A}^L$ . To understand the term  $\mathcal{Q}(N)$  we first provide the intuition for the case  $N = 2$ . Given the typical atom  $\mathbf{z}_o$ , the atom  $\mathbf{z}_1$  may lie anywhere on its boundary and is the center of a ball  $\mathcal{B}(\mathbf{z}_1, r_1)$ , with  $r_1 \leq r_o$ . Furthermore, the atom  $\mathbf{z}_2$ , which is the second part of the mutually-nearest-neighbours pair, will lie on the boundary of  $\mathcal{B}(\mathbf{z}_1, r_1)$ . However the first integral in (7) does not take into account the fact that, the atom  $\mathbf{z}_2$  should not lie inside the first ball  $\mathcal{B}(\mathbf{z}_o, r_o)$ . This probability should be subtracted from the integral and is shown as  $\mathcal{Q}(2)$ . Obviously  $\mathcal{Q}(1) = 0$ , because such an event can not be encountered with less than three atoms. To calculate the term for  $N = 2$ , we use the bound in [KMN06, pp.10-11]

$$\begin{aligned} \mathcal{Q}(2) &\leq \int_{V_o \geq V_1 \geq 0} \frac{V(\mathcal{B}(\mathbf{z}_1, r_o) \cap \mathcal{B}(\mathbf{z}_o, r_o))}{V(\mathcal{B}(\mathbf{z}_1, r_o))} e^{-V_o} e^{-V_1} dV_1 dV_o \\ &= \int_{V_o \geq V_1 \geq 0} \frac{aV_o}{V_o} e^{-V_o} e^{-V_1} dV_1 dV_o = \frac{\alpha}{2}. \end{aligned} \quad (9)$$

For higher number of  $N$ , given that the previous atoms are not contained in the union of balls up to  $N - 1$ , we calculate the correction term  $q(N)$  for the atom with index  $N$ . However, notice that the integral in (7) is equal to  $\frac{N}{(N+1)!} < \alpha/2$  for  $N \geq 3$ , so that no more terms should be calculated for the lower bound, because the expression in (7) is zero for  $N \geq 3$ .

To understand the upper bound, consider first the set of  $N + 1$  overlapping balls with centers the points of the sequence  $\{\mathbf{z}_o, \mathbf{z}_1, \dots, \mathbf{z}_N\}$  and radii  $\{r_o, r_1, \dots, r_N\}$ . The total area  $\mathcal{A}$  depends on the relative positions of the atoms. It is minimum, when the positions are such that the area of overlapping is maximum. This happens when no three or more balls overlap. If - say - three balls overlap, we can always move the center of the third ball but keeping the radius fixed, so that the area of overlap between the third and the second (or the first) increases. Hence, the minimum area  $\mathcal{A}$  results when all atoms fall on a line and have distance  $r_n$ ,  $n = 0, \dots, N$  between them. In such a linear constellation, consider the ball  $\mathcal{B}_o := \mathcal{B}(\mathbf{z}_o, r_o)$ . This ball overlaps with  $\mathcal{B}_1$  and touches  $\mathbf{z}_1$  on its boundary. Obviously, the area of overlap cannot be more than  $V_1/2$ , leaving exactly the same area  $V_1/2$  untouched. Iterating the process, we get  $\mathcal{A}^U = V_o + V_1/2 + \dots + V_{N-1}/2 + V_N/2 \subset \mathcal{A}$ , but considering that  $V_{N-1} = V_N$  we reach the presented bound. The integrals can be easily solved by hand and derive the presented results.

We could simply calculate these nested integrals and be through. However, there is an easier way to find their closed form solution with the following probabilistic interpretation, inspired by [KMN06, p.10]. For the lower bound, the integral (7) is equal to  $1/2$  times ( $V_{N-1} = V_N$ ) the probability that  $V_o \geq \dots \geq V_{N-1} \geq 0$ , where the  $V_i$ 's are independent,  $V_o, \dots, V_{N-2}$  are exponentially distributed and  $V_{N-1}$  is also exponentially distributed with

parameter 2. The probability that  $V_o, \dots, V_{N-2}$  are ordered in this way is  $1/(N-1)!$ . Since  $V_{N-1}$  has the same distribution as the minimum of two independent exponentially distributed random variables, the probability that one of these two will be the smallest among the  $N+1$  (previous  $N-1$  plus these 2) is  $2/(N+1)$ . It follows that the first integral is equal to

$$\frac{1}{2} \frac{1}{(N-1)!} \frac{2}{N+1} = \frac{N}{(N+1)!}.$$

For the upper bound, we make the transformation  $V_i = 2U_i$  in the integral, which results in a term  $2^N$  outside the integral. Consider first the case  $N \geq 3$ . All  $N$  variables are independent exponentially distributed, but  $U_o$  and  $U_{N-1}$  have parameter 2 and can be considered each as the minimum of two independent exponentially distributed random variables with parameter 1. Suppose  $U_{o,a}$  and  $U_{o,b}$  are the auxiliary variables, so that  $U_o = \min\{U_{o,a}, U_{o,b}\}$  and suppose the same for  $U_{N-1}$ . The probability in question is  $1/2$  (case  $U_{o,a} = U_{o,b}$ ) times  $1/2$  (case  $U_{N-1,a} = U_{N-1,b}$ ) the probability that  $U_{o,a}, U_{o,b} \geq U_1 \geq \dots \geq U_{N-2}$  are ordered this way. The latter is  $2/N!$ , where the 2 comes since we are indifferent regarding the ordering of  $U_{o,a}, U_{o,b}$ . Furthermore, the probability that one of the two  $U_{N-1,a}$  or  $U_{N-1,b}$  is the smallest among the  $N+2$  (previous  $N-2$  plus these 2 plus the 2 referring to  $U_o$ ) is equal to  $2/(N+2)$ . Altogether, we get

$$\frac{1}{2} \frac{1}{2} 2^N \frac{2}{N!} \frac{2}{N+2} = 2^N \frac{N+1}{(N+2)!}, \quad N \geq 3.$$

For the case  $N = 2$ ,  $V_o$  and  $V_1$  are both independent exponentially distributed with parameter 1. The probability that  $V_o \geq V_1$  is simply  $1/2$ . For  $N = 1$ , the area  $\mathcal{A}^{UB}(1) = V_o + \frac{1}{2}V_o$ .  $\square$

Observe that the integrals in (7) and (8) consider as variables the volumes  $V_i$ , with unit density, hence  $V_i = \pi r_i^2$ . In the case of  $\lambda \neq 1$ ,  $V_i = \lambda \pi r_i^2$  and the results will be exactly the same, hence the bounds are density-invariant, exactly as stated for the model NN in **PI**. We next present a numerical evaluation of the bounds in Table III and plot the results for comparison in Fig. 9. The table and figure, include values of the generation number probability from Monte Carlo simulations. The area simulated has dimensions  $10 \times 10 [m^2]$  and the density of the simulated p.p.p. is 2 [atoms/ $m^2$ ] hence in average  $\mathbb{E}[N_t] = 200$  atoms are randomly placed per realization. Finally, the values of the generation number probability result after 5000 iterations.

The procedure creates a  $2 \times N_t$  table per iteration ( $N_t$  atoms). The atoms are enumerated in the first row and the first neighbour of each is given in the second. Starting from the first column, the table is searched until the root of the first cluster is found. Once this is done the generation number is equal to the number of hops until the root. The procedure continues to the search starting from the second entry and so on.

We can also check that the bounds derived in Theorem 3 are consistent with the result of Theorem 2, since  $2/3 \geq 0.6215 \geq 1/2$ . Also check how the Monte Carlo value  $0.6207 \approx 0.6215$ .

An alternative way to interpret the generation number probability  $g(k)$  is the following.

TABLE I

NUMERICAL VALUES FOR THE BOUNDS AND MONTE CARLO SIMULATION OF THE GENERATION NUMBER PROBABILITY  $g(k)$ , FOR  $k = 1, \dots, 6$ .

	$g(1)$	$g(2)$	$g(3)$	$g(4)$	$g(5)$	$g(6)$
Upper Bound	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{4}{15}$	$\frac{1}{9}$	$\frac{4}{105}$	$\frac{1}{90}$
Monte Carlo	0.6207	0.2756	0.0815	0.0183	0.0033	0.0006
Lower Bound	$\frac{1}{2}$	0.1378	0	0	0	0

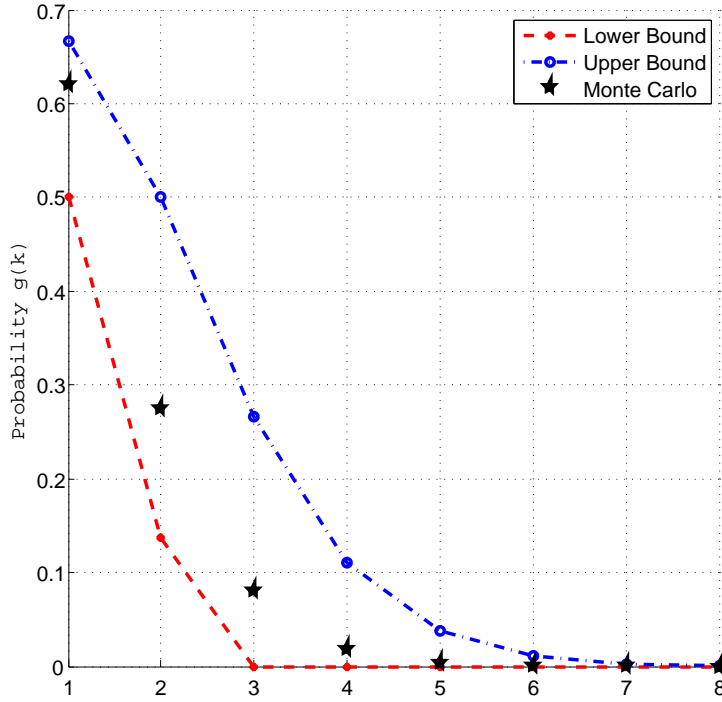


Fig. 9. Evaluation of the upper and lower bounds of the generation number probability  $g(k)$ , for  $k \geq 1$ .

**Corollary 1.** *The event that  $\mathbf{z}_o$  has generation number  $k$  implies that  $\mathbf{z}_o$  belongs to a cluster (typical cluster  $\mathcal{C}_o$ ) of cardinality at least  $k + 1$ , but not the other way round. Consequently, the following relation holds*

$$\mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) \geq k + 1] \geq g(k) \geq \frac{k}{(k+1)!} - \mathcal{Q}(k), \quad k \geq 1. \quad (10)$$

## VI. THE ANCESTOR NUMBER

The generation number refers to exactly this branch of the cluster (tree), where the typical point is included. On each atom of the sequence  $\{\mathbf{z}_o, \mathbf{z}_1, \dots, \mathbf{z}_N\}$ , other branches could be attached, whose number is limited by the maximum number of atoms having the same atom as nearest neighbour (often called *kissing number* [KMN06], [Zon98]). In this sense, the generation number is appropriate for describing the directed graph  $\vec{\mathcal{G}}_{NN}$ , but is not necessarily a good indicator of the size of the typical cluster of  $\mathcal{G}_{NN}$ , so other measures may be more appropriate.

### A. Definition and Bounds

**Definition 1.** Consider an atom of the p.p.p.  $\mathbf{z}_n$  and suppose it belongs to cluster  $\mathcal{C}_o$ . Then, the **ancestor number**  $k_a$  is defined iteratively:

- 1) The two atoms of the mutually-nearest-neighbour pair of  $\mathcal{C}_o$  (see Lemma 1) are given  $k_a = 1$  and form the set  $\mathcal{S}_1$ . By convention  $\mathcal{S}_o$  contains just one of the two atoms of the pair, and initiates the process.
- 2) The atom  $\mathbf{z}_j \in \mathcal{C}_o$  which is closest to the set  $\mathcal{S}_1$ , is given  $k_a = 2$ . Include the new atom in  $\mathcal{S}_2 = \mathcal{S}_1 \cup \{\mathbf{z}_j\}$ .
- 3) Find the closest atom in  $\mathcal{C}_o$  to  $\mathcal{S}_2$ , which is given ancestor number  $k_a = 3$ . Include it in the set  $\mathcal{S}_3$ .
- 4) ...

In this way all  $\mathbf{z}_n \in \mathcal{C}_o$  receive an ancestor number.

Obviously, the ancestor number is increasing with the distance of the atom from the origin pair. By definition, an atom of the p.p.p. receives *ancestor number*  $k_a = K$ , if and only if it is the closest neighbour of a set  $\mathcal{S}_K$ , and it itself has one of the members of  $\mathcal{S}_K$  as nearest neighbour. When the latter is not true, the atom is not included in the cluster and  $\text{card}(\mathcal{C}_o) = K$ . In this sense, the ancestor number best describes the cluster size, in terms of the count of its members. The furthest atom of a cluster to the origin pair, takes ancestor number equal to the cluster size.

Let us now relate the ancestor number to the *generation number* by means of an example. Assume that the cluster  $\mathcal{C}_o$ , has  $n$  branches stemming from the pair and hence,  $n$  atoms with generation number 2. Among these  $n$ , the atom with the shortest distance from (one of the two atoms of) the pair receives  $k_a = 2$ . The ancestor number  $k_a = 3$  is given to the atom with minimum distance from the new set  $\mathcal{S}_2$ , which includes the two atoms with  $k_a = 1$  and the one with  $k_a = 2$ . Observe that, this may be either one of the  $n - 1$  remaining atoms with generation number 2, or the atom with generation number 3, following the atom with  $k_a = 2$  on its branch. This explanation emphasizes the differences between the two numbers (*generation* and *ancestor*).

Based on the ancestor number, given a realization  $\phi$  of a p.p.p. we can provide an **algorithm** to form NN clusters of increasing size:

- 1) Identify the mutually-nearest-neighbour pairs of the realization. These constitute the "roots" of all clusters  $\mathcal{C}_m$  in the network and have  $k_a = 1$ . This is done by finding the pairs of atoms  $\mathbf{z}_i, \mathbf{z}_j$  with the property that the ball  $\mathcal{B}(\mathbf{z}_i, R_1)$ , with  $R_1 = d_{ij} := |\mathbf{z}_i - \mathbf{z}_j|$  is empty and has atom  $j$  on its boundary and the other way round.
- 2) Let the balls  $\mathcal{B}_i$  and  $\mathcal{B}_j$  grow further, with common radius, until their union meets an atom on its boundary, say  $\mathbf{z}_k$ . The radius is now  $R_2 = \min\{d_{ik}, d_{jk}\}$ . This candidate to be included in the cluster with  $k_a = 2$ , should also have one of the two atoms in  $\mathcal{S}_1$  as its nearest neighbour and for this, the ball  $\mathcal{B}(\mathbf{z}_k, R_2)$  should be empty. If true we continue, else we stop and  $\text{card}(\mathcal{C}_o) = 2$ .
- 3) Let the balls  $\mathcal{B}_i, \mathcal{B}_j, \mathcal{B}_k$  grow further, until their union meets an atom on its boundary, say  $\mathbf{z}_l$ . The radius is

$R_3 = \min \{d_{il}, d_{jl}, d_{kl}\}$ . This atom takes the ancestor number  $k_a = 3$ , if one of the three atoms in  $\mathcal{S}_2$  are its nearest neighbour, in other words, the ball  $\mathcal{B}(\mathbf{z}_l, R_3)$  is empty. etc.

4) ...

How the algorithm gradually finds the elements of a cluster until ancestor number  $k_a = 3$ , are illustrated in Fig.10(a)-10(d). By letting the radius  $R$  of the balls grow large enough, the algorithm will eventually converge to the  $\mathcal{G}_{NN}$  graph. For a cluster with  $\text{card}(\mathcal{C}_o) = K + 1$ , an empty ball of **the same** radius  $R_K$  is centered on each of its atoms. Hence the probability of such cluster to appear, is equal to the probability that this union of  $K + 1$  balls is empty

$$\mathcal{A} := \bigcup_{k_a=1,1,2,\dots,K} \mathcal{B}(\mathbf{z}_{k_a}, R_K) = \emptyset. \quad (11)$$

Finally we note here, that the way the ancestor number is defined, gives rise to a sequence of *ancestor radii*  $\{R_1, \dots, R_K\}$  for a cluster of size  $K + 1$ . Each distance  $R_n$  is the minimum distance of the atom with ancestor number  $k_a = n$ , to the set of its  $n$  ancestors  $\mathcal{S}_{n-1}$ .

The probability that a typical point  $\mathbf{z}_o$  has ancestor number  $k_a$  is denoted by  $g_a(k_a)$ . In order to derive certain bounds, we need the following lemmas. Let us denote by *root*, one of the two atoms of the mutually-nearest-neighbour pair. We will make use of the *nearest-neighbour-to-root radii*  $\{\tilde{r}_1, \dots, \tilde{r}_K\}$  which is the sequence of distances of the  $n$ -th nearest neighbour to the root with  $1 \leq n \leq K$ , the sequence of *ancestor radii*  $\{R_1, \dots, R_K\}$  defined above, and the sequence of *auxiliary radii*  $\{R_1, \dots, R_K\}$  to be defined in Lemma 4. We remind the reader here, that a r.v.  $Y$  stochastically dominates  $X$ , and we write  $X \preceq_{st} Y$  if  $\mathbb{P}[X > t] \leq \mathbb{P}[Y > t]$ , for all  $t$ .

**Lemma 3.** *The vector of  $K$  nearest-neighbour-to-root radii  $\{\tilde{r}_1, \dots, \tilde{r}_K\}$  stochastically dominates the vector of  $K$  ancestor radii  $\mathbf{R} = (R_1, \dots, R_K)$  in the sense that,*

$$R_{k_a} \preceq_{st} \tilde{r}_{k_a}, \quad 1 \leq k_a \leq K. \quad (12)$$

*The joint probability distribution (p.d.f.) of the distances of  $K$  nearest neighbours to the root, given that  $\tilde{V}_k = \lambda\pi\tilde{r}_k^2$ ,  $k \in \{1, \dots, K\}$  is*

$$f_{\tilde{V}_1, \dots, \tilde{V}_K}(\tilde{V}_1, \dots, \tilde{V}_K) = e^{-\tilde{V}_K}, \quad \tilde{V}_K \geq \dots \geq \tilde{V}_1 \geq 0. \quad (13)$$

*Proof.* The domination result, can be derived, by first exploring the case  $k_a = 2$  and applying similar arguments to higher ancestor numbers. Suppose the mutually-nearest-neighbour pair consists of atoms  $\mathbf{z}_i, \mathbf{z}_j$ . Obviously  $R_1 = \tilde{r}_1$ . From the second step of the algorithm, the two balls around these atoms should expand their volume with common radius, until they touch  $\mathbf{z}_k$ , which is the candidate for  $k_a = 2$ . This will occur at a distance  $R_2 \leq \tilde{r}_2$ , because  $\mathbf{z}_k$ , might be the second neighbour of  $\mathbf{z}_j$  instead of the root atom  $\mathbf{z}_i$ . In this sense, the probability that  $\mathbf{z}_k$



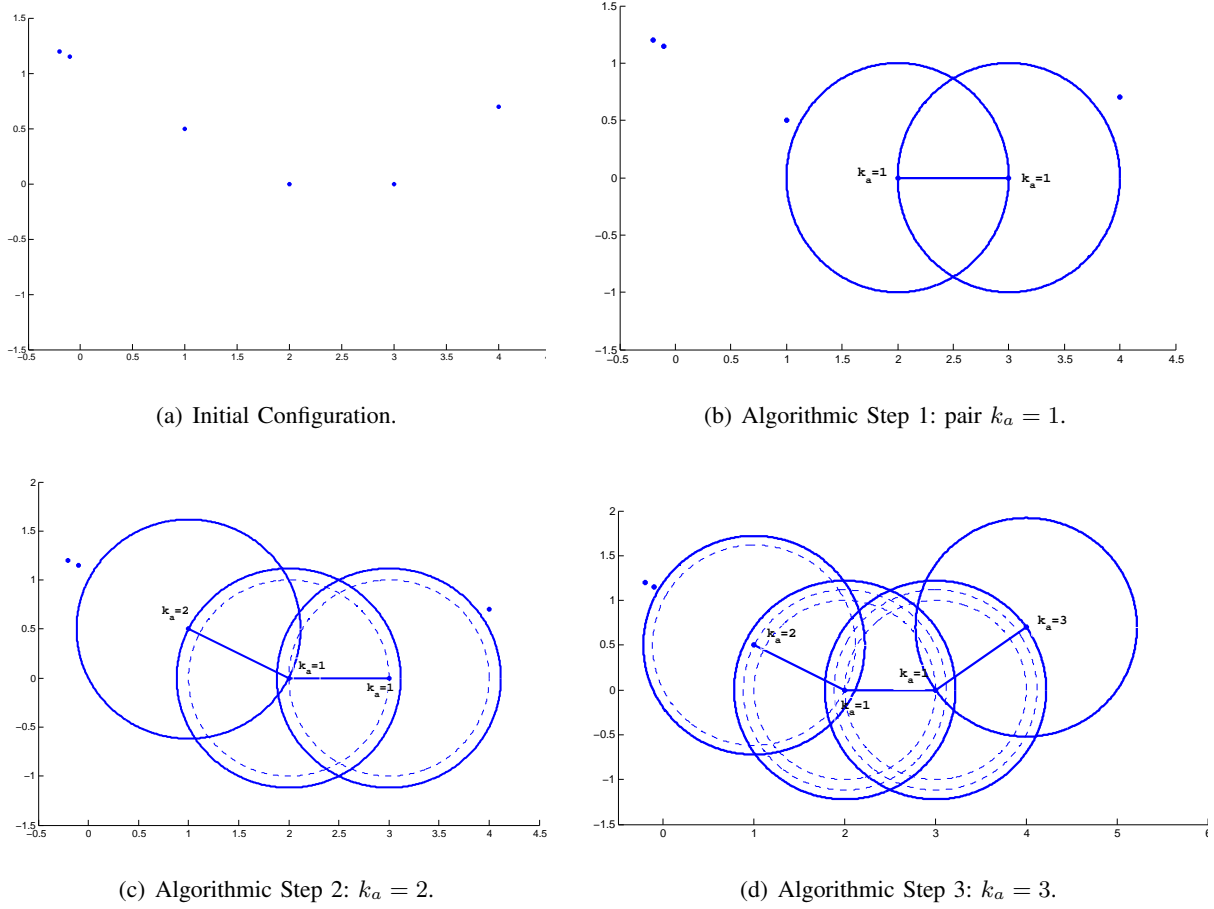


Fig. 10. Implementation of the algorithm to form NN clusters based on the ancestor number. In the example, the algorithm forms a cluster of 4 elements (atoms). The dashed circles in subfigures (c) and (d) illustrate the size of the circles in the previous steps of the algorithm. In subfigure (d) we can identify 2 different branches connected to the origin pair, each one of which has a depth of 1 atom. By further enlarging the area of (d) the candidate for  $k_a = 4$  would be met, but it is in nearest neighbour relation with a different atom than those of the cluster (the one exactly at its left). This is why the algorithm stops at  $k_a = 3$  and the cluster has cardinality 4.

is met at  $R_2 = \tilde{r}_2 = t$ , when  $R_1 = \tilde{r}_1 = s$  can be calculated for both cases to be  $e^{-V(\mathcal{B}_i(t) \cup \mathcal{B}_j(t)) + V(\mathcal{B}_i(s) \cup \mathcal{B}_j(s))} < e^{-V(\mathcal{B}_i(t)) + V(\mathcal{B}_j(s))}$  and as a result  $\tilde{r}_2$  stochastically dominates  $R_2$ .

To further derive the p.d.f. for each  $n$ , we should first calculate the conditional probability that  $\tilde{V}_K \leq t$ , given that  $\tilde{V}_{K-1} = s$ , with  $t \geq s$ . In the p.p.p. case, this is equal to

$$\begin{aligned} \mathbb{P} \left[ \tilde{V}_K \leq t | \tilde{V}_{K-1} = s \right] &= 1 - e^{-(t-s)} \Rightarrow \\ f_{\tilde{V}_K | \tilde{V}_{K-1}}(t|s) &= e^{-(t-s)} \end{aligned} \quad (14)$$

Obviously by unconditioning iteratively from  $\tilde{V}_{K-1}$  to  $\tilde{V}_1$  and using (14) with appropriate replacement of the index, we reach the result. The case for  $K = 2$  has already been shown and used in [BG13] and [GB13].  $\square$

**Lemma 4.** *The vector of  $K$  ancestor radii  $\mathbf{R} = (R_1, \dots, R_K)$  stochastically dominates the auxiliary  $K$ -length vector  $\mathbf{R}' = (R'_1, \dots, R'_K)$  in the sense that,*

$$R'_{k_a} \preceq_{st} R_{k_a}, \quad 1 \leq k_a \leq K, \quad (15)$$

*The vector of auxiliary random variables (r.v.'s)  $\mathbf{R}'$  has joint p.d.f. (with  $V'_k = \lambda\pi R'^2_k$ )*

$$f_{V'_1, \dots, V'_K}(V'_1, \dots, V'_K) = K! e^{-KV'_K} e^{V'_{K-1}} \dots e^{V'_1}, \quad V'_K \geq \dots \geq V'_1 \geq 0. \quad (16)$$

*Proof.* The auxiliary vector  $\mathbf{R}'$  is defined iteratively as follows. Let the first element  $R'_1$  be the distance between the two atoms  $\mathbf{z}_i$  (root),  $\mathbf{z}_j$  of the pair, exactly as the ancestor distance, and  $R'_1 := R_1$ . For analytical purposes also  $R'_0 := 0$ . Ancestor distance  $R_2$  is explained in Step 2 of the previous algorithm to be the closest distance of an atom  $\mathbf{z}_k$  to one of the two atoms  $\mathbf{z}_i, \mathbf{z}_j$ . To find this geometrically, the balls  $\mathcal{B}_i, \mathcal{B}_j$  should grow with common radius until they meet  $\mathbf{z}_k$  on their common boundary. Thus,  $\mathbf{z}_k$  is the candidate for the ancestor number  $k_a = 2$ . The probability that  $R_2 > t$ , given that  $R_1 = s$  is equal to  $e^{-V(\mathcal{B}_i(t) \cup \mathcal{B}_j(t)) + V(\mathcal{B}_i(s) \cup \mathcal{B}_j(s))}$  for the p.p.p. case. Let the auxiliary random variable  $R'_2$  be the common radius of the balls  $\mathcal{B}_i, \mathcal{B}_j$  when one of them meets  $\mathbf{z}_k$  on its boundary, but in this case the one ball is "far away" from the other, hence there is no overlap. Obviously, this probability  $e^{-V(\mathcal{B}_i(t)) - V(\mathcal{B}_j(t)) + V(\mathcal{B}_i(s)) + V(\mathcal{B}_j(s))} < e^{-V(\mathcal{B}_i(t) \cup \mathcal{B}_j(t)) + V(\mathcal{B}_i(s) \cup \mathcal{B}_j(s))}$  when  $R_2 = R'_2 = t$  and  $R_1 = R'_1 = s$ , hence  $R'_2$  is stochastically dominated by  $R_2$ . Repeating this argument stepwise for  $3, \dots, K$ , we reach the domination result in (15).

We enumerate the atoms by their ancestor number, giving to the root the number 0. Let  $V'_{n,K}$ ,  $n = 0, \dots, K-1$  be the volume of the ball centered at  $\mathbf{z}_n$ , when the atom with ancestor number  $K$  is touched by one of the  $K$  balls (ancestors with balls  $\mathcal{B}_0, \dots, \mathcal{B}_{K-1}$ ). To find the joint p.d.f. of  $\mathbf{R}'$  we first need to write the conditional probability

$$\begin{aligned} & \mathbb{P}[\min\{V'_{0,K}, \dots, V'_{K-1,K}\} \leq t | V'_{0,K-1} = \dots = V'_{K-1,K-1} = s \leq t] \\ & \stackrel{(a)}{=} 1 - (\mathbb{P}[V'_{0,K} > t | V'_{0,K-1} = s \leq t])^K \\ & \stackrel{p.p.p. \text{ def}}{=} 1 - (e^{-t+s})^K, \quad K \geq 1. \end{aligned} \quad (17)$$

To better clarify the formula, we give here two examples. For  $K = 1$ ,  $\mathbb{P}[V'_{0,1} \leq t | V_0 = 0] = 1 - e^{-t}$  because per definition (above)  $R'_0 = 0$ . For  $K = 2$ ,  $\mathbb{P}[\min\{V'_{0,2}, V'_{1,2}\} \leq t | V'_{0,1} = V'_{1,1} = s] = 1 - e^{-2t+2s}$  (see also Fig.10(c) for an illustrative explanation). In the equations, (a) comes from the fact that the  $K$  events of a ball centered at  $\mathbf{z}_n$  with radius  $R'_K$  having an empty ring in  $[R'_K, R'_{K-1}]$  (i.e. the volume  $V'_{n,K} - V'_{n,K-1}$ ,  $n = 0, \dots, K-1$  is empty of atoms) are i.i.d. because the p.p.p. is homogenous and the balls are considered "far away" from each other. The joint p.d.f. in (16) results by first differentiating over  $t$

$$f_{V'_{0,K} | V'_{0,K-1}} = K e^{-Kt+Ks}, \quad (18)$$

and further applying Bayes' rule with the product

$$\prod_{n=1}^K f_{V_{0,n}|V_{0,n-1}} = \prod_{n=1}^K n e^{-ns_n + ns_{n-1}}, \quad s_0 = 0.$$

□

**Theorem 4.** *Given a planar p.p.p.  $\Phi$  the probability  $g_a(k_a)$  that the typical atom  $\mathbf{z}_o$  has ancestor number  $k_a$  can be bounded as follows*

$$\frac{k_a!}{\prod_{n=0}^{k_a-1} (k_a + 1 - \alpha - n(1 - \frac{1}{\pi}))} \geq g_a(k_a) \geq \frac{1}{(1 + k_a(1 - \alpha))^{k_a}} - \tilde{Q}(k_a), \quad k_a \geq 1. \quad (19)$$

The correction term  $\tilde{Q}(k_a) := \tilde{q}(k_a) + \sum_{i=1}^{k_a-1} \tilde{Q}(i) - \tilde{Q}(i-1)$ ,  $\tilde{Q}(0) = 0$ ,  $\tilde{Q}(k_a) \geq 0$ , is increasing in  $k_a$ . It holds  $\tilde{Q}(1) = \tilde{q}(1) = 0$  and  $\tilde{Q}(k_a) \geq \tilde{Q}(2) = 0.0795$ . The lower bound expression is equal to zero for  $k_a \geq 3$ . The constant  $\alpha = \frac{2}{3} - \frac{\sqrt{3}}{2\pi}$ .

*Proof.* For the lower bound we first choose some radius  $R_{k_a} = t$ . Then the volume of the area  $\mathcal{A}^{LB}(t)$

$$V(\mathcal{A}^{LB}(t)) = k_a(1 - \alpha)V_0(t) \geq V\left(\bigcup_{n=1}^{k_a} \mathcal{B}_n(t) \setminus \mathcal{B}_{n-1}(t)\right) \stackrel{(11)}{=} V(\mathcal{A} \setminus \mathcal{B}_0(t)),$$

because the distances between atoms of the cluster are at most  $t$ . We further make use of Lemma 3 which states that  $R_{k_a} \preceq_{st} \tilde{r}_{k_a}$ . As a result the area  $\mathcal{A}^{LB}$  with radius  $\tilde{r}_{k_a}$  is stochastically larger than the same area with radius  $R_{k_a}$ . The bound is further calculated using the p.d.f. in (13)

$$\int_0^\infty d\tilde{V}_1 \cdots \int_{\tilde{V}_{k_a-2}}^\infty d\tilde{V}_{k_a-1} \int_{\tilde{V}_{k_a-1}}^\infty d\tilde{V}_{k_a} e^{-\tilde{V}_{k_a}} e^{-k_a(1-\alpha)\tilde{V}_{k_a}} - \tilde{Q}(k_a). \quad (20)$$

In the above,  $\tilde{Q}(k_a)$  is a correction term, which is non-decreasing in  $k_a$ . To understand this, consider the case  $k_a = 2$  in Fig. 10(c). Suppose that the atom with  $k_a = 2$  is the typical atom  $\mathbf{z}_2 := \mathbf{z}_o$ , with distance  $R_2$  from its first neighbour ( $\mathbf{z}_1$ ) with  $k_a = 1$  (the closest of the two roots). The ball  $\mathcal{B}(\mathbf{z}_o, R_2)$  is empty and the atom  $\mathbf{z}_1$  lies on its boundary. However this is in mutually-nearest-neighbour relation with  $\mathbf{z}_0$ , which lies on a distance  $R_1$  at the boundary of the ball  $\mathcal{B}(\mathbf{z}_1, R_1)$ . It should necessarily lie outside the ball  $\mathcal{B}(\mathbf{z}_o, R_2)$ , so the integral on the left-hand side of (20) should not consider such points for  $\mathbf{z}_0$ . Having said this, the correction term can be calculated using the inequality

$$\begin{aligned} \int_{\mathbf{z}_0 \in \mathcal{B}(\mathbf{z}_2, R_2) \cap \mathcal{B}(\mathbf{z}_0, R_1)} f(\mathbf{z}_0) d\mathbf{z}_0 &\leq \frac{V(\mathcal{B}(\mathbf{z}_2, R_2) \cap \mathcal{B}(\mathbf{z}_0, R_1))}{V(\mathcal{B}(\mathbf{z}_2, R_2))} \int_{\mathbf{z}_0 \in \mathcal{B}(\mathbf{z}_2, R_2)} f(\mathbf{z}_0) d\mathbf{z}_0 \\ &\leq \alpha \int_{\mathbf{z}_0 \in \mathcal{B}(\mathbf{z}_2, R_2)} f(\mathbf{z}_0) d\mathbf{z}_0. \end{aligned} \quad (21)$$

From the above, the correction term for  $k_a = 2$  is

$$\begin{aligned}
\tilde{Q}(2) = \tilde{q}(2) &\leq \alpha \int_0^\infty d\tilde{V}_1 \int_{\tilde{V}_1}^\infty d\tilde{V}_2 e^{-\tilde{V}_2} e^{-2(1-\alpha)\tilde{V}_2} \\
&= \frac{\alpha}{(1+2(1-\alpha))^2} = 0.0795.
\end{aligned} \tag{22}$$

The correction for  $k_a = 3$  will use this term and add the extra  $\tilde{q}(3)$ , which comes for the case that the point  $\mathbf{z}_0$  falls in the union of balls  $\mathcal{B}(\mathbf{z}_3, R_3) \cup \mathcal{B}(\mathbf{z}_2, R_2)$ , given that  $\mathbf{z}_1$ , does not fall in the ball  $\mathcal{B}(\mathbf{z}_3, R_3)$  ( $= \tilde{q}(2)$ ). This means that the correction term  $\tilde{Q}(k_a) \geq \tilde{Q}(2)$ , for  $k_a \geq 2$ . As a result the lower bound is 0 for  $k_a \geq 3$ .

For the upper bound we again choose some radius  $R_{k_a} = t$ . Then the volume of the area  $\mathcal{A}^{UB}(t)$

$$V(\mathcal{A}^{UB}(t)) = (1-\alpha)V_0(t) \leq V\left(\bigcup_{n=1}^{k_a} \mathcal{B}_n(t) \setminus \mathcal{B}_{n-1}(t)\right) \stackrel{(11)}{=} V(\mathcal{A} \setminus \mathcal{B}_0(t)),$$

because the area at the left-hand side is exactly the area  $\mathcal{B}_{k_a-1}(t) \setminus \mathcal{B}_{k_a}(t)$  (equal to the symmetric  $\mathcal{B}_{k_a}(t) \setminus \mathcal{B}_{k_a-1}(t)$ ) and does not contain any part of the rest  $k_a - 1$  planar subsets  $\mathcal{B}_n(t) \setminus \mathcal{B}_{n-1}(t)$ . Including part of these further tightens the bound. Observe that all the balls have the same radius  $t$  and distances between them equal to  $R_n \leq t$ ,  $1 \leq n \leq k_a$ . Obviously, when the ancestor number is  $k_a$ , then  $t = R_{k_a}$ . The area of  $\mathcal{B}_n(t) \setminus \mathcal{B}_{n-1}(t)$  can be given as a result of the area of overlap of two circles with the same radius  $t = R_{k_a}$  and distance between centers  $R_n$ . The area in question is equal to [Wei]

$$\mathcal{B}_n(R_{k_a}) \setminus \mathcal{B}_{n-1}(R_{k_a}) = \pi R_{k_a}^2 \left(1 - \frac{2}{\pi} \arccos\left(\frac{R_n}{2R_{k_a}}\right) + \frac{R_n}{2\pi R_{k_a}} \sqrt{4 - \frac{R_n^2}{R_{k_a}^2}}\right). \tag{23}$$

Observe that for  $R_n = R_{k_a}$ , the expression above reduces to  $\pi R_{k_a}^2 (1-\alpha)$ . We use the first order Maclaurin series approximation of the arccos function to bound the area, which gives  $\arccos(x) \leq \frac{\pi}{2} - x$ . As a result, the total area

$$\begin{aligned}
\sum_{n=1}^{k_a} \mathcal{B}_n(R_{k_a}) \setminus \mathcal{B}_{n-1}(R_{k_a}) &= V(\mathcal{B}_{k_a}(R_{k_a}) \setminus \mathcal{B}_{k_a-1}(R_{k_a})) + \sum_{n=1}^{k_a-1} V(\mathcal{B}_n(R_{k_a}) \setminus \mathcal{B}_{n-1}(R_{k_a})) \\
&\geq \pi R_{k_a}^2 \left( (1-\alpha) + \sum_{n=1}^{k_a-1} \left(1 - \frac{2}{\pi} \left(\frac{\pi}{2} - \frac{R_n}{2R_{k_a}}\right) + \frac{R_n}{2\pi R_{k_a}} \sqrt{4 - \frac{R_n^2}{R_{k_a}^2}}\right) \right) \\
&\geq \pi R_{k_a}^2 \left( (1-\alpha) + \sum_{n=1}^{k_a-1} \frac{R_n}{\pi R_{k_a}} \right) \\
&\stackrel{(a)}{=} V_{k_a} \left( (1-\alpha) + \sum_{n=1}^{k_a-1} \frac{\sqrt{V_n}}{\pi \sqrt{V_{k_a}}} \right) \\
&\stackrel{(b)}{\geq} V_{k_a} (1-\alpha) + \sum_{n=1}^{k_a-1} \frac{V_n}{\pi} = V(\mathcal{A}^{UB_2}(R_{k_a})).
\end{aligned}$$

The equality (a) comes by change of variables and the inequality (b) results from the fact that  $V_{k_a} \geq V_n$ . After bounding the area, we use Lemma 4 which states that,  $R_{k_a} \preceq_{st} R_{k_a}$ . As a result, the area  $\mathcal{A}^{UB_2}$  with radius

$R_{k_a}$  is stochastically smaller than the same area with radius  $R_{k_a}$ . The upper bound is further calculated using the p.d.f. in (16)

$$k_a! \int_0^\infty dV'_1 e^{V'_1} \dots \int_{V'_{k_a-2}}^\infty dV'_{k_a-1} e^{V'_{k_a-1}} \int_{V'_{k_a-1}}^\infty dV'_{k_a} e^{-k_a V'_{k_a}} e^{-(1-\alpha)V'_{k_a}} e^{-\sum_{n=1}^{k_a-1} \frac{V'_n}{\pi}} =$$

$$k_a! \int_0^\infty dV'_1 e^{(1-\frac{1}{\pi})V'_1} \dots \int_{V'_{k_a-2}}^\infty dV'_{k_a-1} e^{(1-\frac{1}{\pi})V'_{k_a-1}} \int_{V'_{k_a-1}}^\infty dV'_{k_a} e^{-k_a V'_{k_a}} e^{-(1-\alpha)V'_{k_a}}.$$

□

As in the case of the generation number, the bounds for the ancestor number probability are density-invariant as well (see **P.I**). We next give a numerical evaluation of the bounds in Table II and plot the results for comparison in Fig. 11.

TABLE II

NUMERICAL VALUES FOR THE BOUNDS OF THE ANCESTOR NUMBER PROBABILITY  $g_a(k_a)$ , FOR  $k_a = 1, \dots, 6$ .

	$g_a(1)$	$g_a(2)$	$g_a(3)$	$g_a(4)$	$g_a(5)$	$g_a(6)$
Upper Bound	0.6215	0.3977	0.2529	0.1593	0.0996	0.0618
Lower Bound	0.6215	0.1238	0	0	0	0

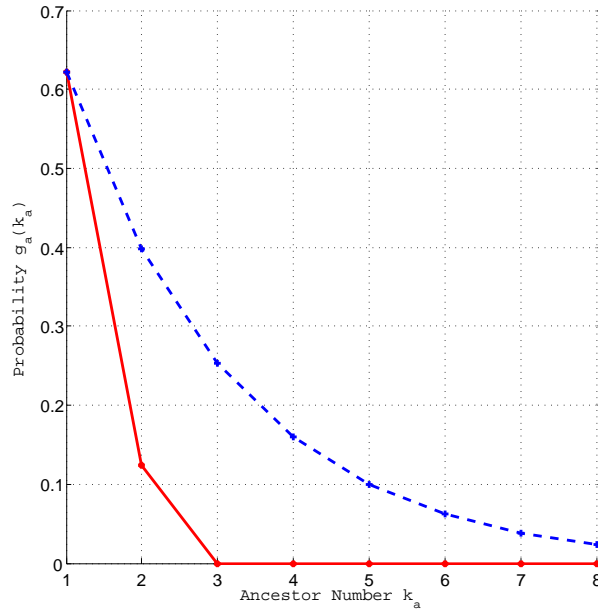


Fig. 11. Evaluation of the upper and lower bounds of the ancestor number probability  $g_a(k_a)$ , for  $k_a \geq 1$ .

### B. Interpretation of the ancestor number: Cluster size distribution

The ancestor number is, as explained at the beginning of the previous paragraph, a more precise number to describe the cluster size, compared to the generation number. A first important observation is that the set of events  $\mathcal{E}_{k_a} = \{\text{the typical atom has ancestor number } k_a\}$  exhausts the event space and hence,

$$\sum_{k_a=1}^{\infty} g_a(k_a) = 1. \quad (24)$$

To get a better understanding of the ancestor number, consider a finite set of atoms, with size  $N$ . The number of atoms will tend to infinity  $N \rightarrow \infty$  as we approach the actual realization of the p.p.p.  $\Phi$ . Let the algorithm described in the previous paragraph evolve stepwise, and first find all atoms with ancestor number 1 (in step 1). These are  $N_1$  out of the total  $N$  atoms and constitute the "roots" of all clusters with higher size. In other words, the total number of existing clusters in the finite setting is  $N_1/2$  (because both mutually-nearest-neighbours take ancestor number 1). A subset of these will add (in Step 2) another atom with ancestor number 2. The number of clusters with cardinality at least 3 will be  $N_2 \leq N_1/2$ . The process will continue in the next steps of the algorithm, resulting in the inequalities

$$\frac{N_1}{2} \geq N_2 \geq N_3 \geq \dots \quad (25)$$

Dividing the above expression by  $N$  and letting  $N \rightarrow \infty$  we have proven that:

**Lemma 5.** *The probability that the ancestor number of the typical atom is  $k_a$ , is ordered as follows*

$$\frac{g_a(1)}{2} \geq g_a(2) \geq g_a(3) \geq \dots \quad (26)$$

This Lemma further tightens the upper bound of the ancestor number probability, as shown in Fig.12.

Using the above line of thought, the probability that the cardinality of the typical cluster is larger or equal to  $k_a + 1$  is exactly defined by the probability of the ancestor number  $k_a$ .

**Theorem 5.** *The probability that the typical atom belongs to a cluster of size  $k_a + 1$ , is fully defined by the probability of the typical atom having ancestor number  $k_a$ , i.e.*

$$\begin{aligned} \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = k_a + 1] &= (k_a + 1)(g_a(k_a) - g_a(k_a + 1)), \quad k_a \geq 2 \\ \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = 2] &= 2(g_a(1)/2 - g_a(2)), \quad k_a = 1. \end{aligned} \quad (27)$$

The tail probability of the above event is equal to

$$\mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) \geq k_a + 1] = k_a g_a(k_a) + \sum_{n=k_a}^{\infty} g_a(n), \quad k_a \geq 2 \quad (28)$$

From (27) and (24),  $\mathbb{P}[\text{card}(\mathcal{C}_o) \geq 2] = \sum_{n=1}^{\infty} g_a(n) = 1$ .

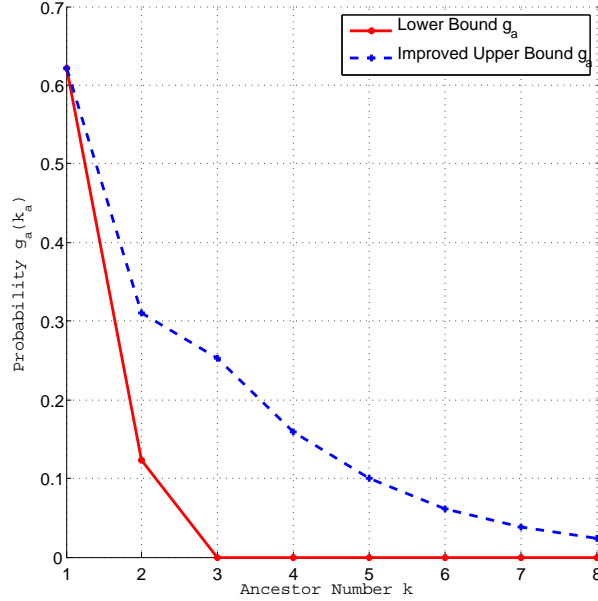


Fig. 12. Tightened upper bound of the ancestor number probability  $g_a(k_a)$ , for  $k_a \geq 1$ .

*Proof.* To prove this, we use the finite model above with  $N$  total atoms and  $N_{k_a}$  atoms with ancestor number  $k_a$  (see also (26)). Suppose the maximum cluster size is  $K > 2$ . Then there are  $N_{K-1}$  atoms with ancestor number  $k_a = K - 1$ . Similarly, there are  $N_{K-2}, \dots, N_1$  atoms with ancestor number  $k_a = K - 2, \dots, 1$ . Since, the size of the maximum cluster is  $K$ , the total atoms which belong to these clusters is  $KN_K$ . Then,

$$\begin{aligned} \frac{KN_{K-1}}{N} &\rightarrow \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = K]. \\ \frac{(K-1)(N_{K-2} - N_{K-1})}{N} &\rightarrow \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = K-1]. \\ &\dots \\ \frac{3(N_2 - N_3)}{N} &\rightarrow \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = 2]. \end{aligned}$$

For the case  $k_a = 1$ , we have a small difference, because both atoms of the root take ancestor number equal to 1. Hence

$$\frac{2(N_1/2 - N_2)}{N} \rightarrow \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = 2].$$

By letting  $N \rightarrow \infty$  the left-hand side gives the probability that the typical atom belongs to a cluster of size  $k_a + 1$ ,

$$\begin{aligned} \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = k_a + 1] &= (k_a + 1)(g_a(k_a) - g_a(k_a + 1)), \quad k_a \geq 2 \\ \mathbb{P}[\mathbf{z}_o \in \mathcal{C}_o, \text{card}(\mathcal{C}_o) = 2] &= 2(g_a(1)/2 - g_a(2)), \quad k_a = 1. \end{aligned} \tag{29}$$

□

The tail distribution of the cluster size, is bounded using the bounds of the ancestor number in Theorem 4 and Lemma 5. It is plotted in Fig.13. Furthermore, using the known bounds for the ancestor number (upper bound  $g_a \leq g_a^{UB}$  for the positive term of the difference and lower bound  $-g_a \leq -g_a^{LB}$  for the negative term, we can find a loose upper bound for the exact cluster size. The most important result from the plots is to illustrate how the probability of a higher cluster size (Fig.14) diminishes as  $k_a + 1$  increases. Observe that the loose upper bound is already 10% for cluster size 11, implying that cluster sizes of average or large size are highly improbable. This is better supported by the tail probability upper bound, which is also around 10% for cluster size 11.

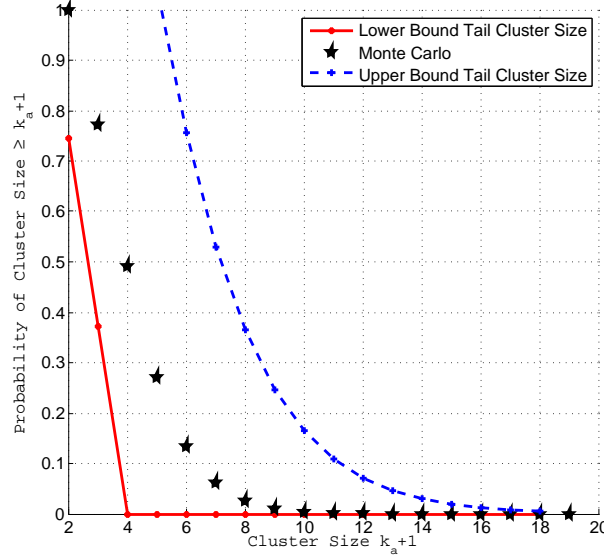


Fig. 13. Distribution of the tail probability for the cluster size  $\mathbb{P}[\text{card}(\mathcal{C}_o) \geq k_a + 1]$ .

The two figures also include the Monte Carlo values of the cluster size probability. These are derived by simulation over an area with dimensions  $10 \times 10 [m^2]$  and a p.p.p. with density 2 [atoms/ $m^2$ ] hence in average  $\mathbb{E}[N_t] = 200$  atoms are randomly placed per realization. The values of the cluster size probability result after 5000 iterations.



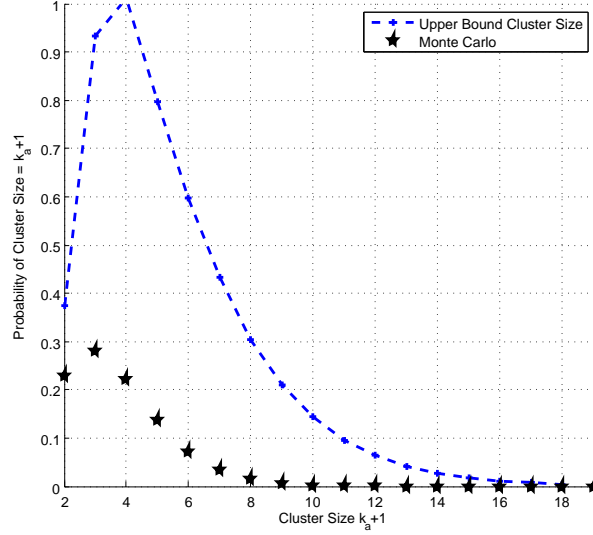


Fig. 14. Upper bound for the probability of the cluster size  $\mathbb{P}[\text{card}(\mathcal{C}_o) = k_a + 1]$ ,  $k_a \geq 2$ .

TABLE III

NUMERICAL VALUES FOR THE BOUNDS AND MONTE CARLO SIMULATION OF THE CLUSTER SIZE AND TAIL PROBABILITY, FOR  $k = 2, \dots, 8$ .

$\text{card}(\mathcal{C}_o) =$	2	3	4	5	6	7	8
Upper Bound	0.3739	0.9323	1.0116	0.7967	0.5974	0.4325	0.3050
Monte Carlo	0.2287	0.2796	0.2219	0.1362	0.0727	0.0347	0.0163
$\text{card}(\mathcal{C}_o) \geq$	2	3	4	5	6	7	8
Upper Bound	1.6037	1.6037	1.4301	1.0558	0.7570	0.5303	0.3647
Monte Carlo	0.9999	0.7712	0.4916	0.2697	0.1335	0.0608	0.0261
Lower Bound	0.7453	0.3714	0	0	0	0	0

## REFERENCES

- [AJ12] S. Akoum and R. W. Heath Jr. Multi-cell coordination: A Stochastic Geometry approach. *13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2012.
- [AJ13] S. Akoum and R. W. Heath Jr. Interference coordination: Random clustering and adaptive limited feedback. *IEEE Trans. on Signal Processing*, 61, no. 7:1822–1834, April 2013.
- [BG13] F. Baccelli and A. Giovanidis. Coverage by pairwise base station cooperation under adaptive geometric policies. *Proc. 47th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, Nov. 2013.
- [BG15] F. Baccelli and A. Giovanidis. A Stochastic Geometry Framework for Analyzing Pairwise-Cooperative Cellular Networks. *IEEE Trans. on Wireless Communications*, 14(2), Feb. 2015.
- [BK15] B. Błaszczyszyn and H.P. Keeler. Studying the SINR process of the typical user in poisson networks by using its factorial

- moment measures. *Information Theory, IEEE Transactions on*, 61(12):6774–6794, 2015.
- [DSS99] D.J. Daley, H. Stoyan, and D. Stoyan. The volume fraction of a Poisson germ model with maximally non-overlapping spherical grains. *Adv. Appl. Prob. (SGSA)*, 31:610–624, 1999.
- [FEC02] R. Fabbri, L.F. Estrozi, and L. Da F. Costa. On Voronoi Diagrams and Medial Axes. *Journal of Mathematical Imaging and Vision*, 17:27–40, 2002.
- [GACD15] A. Giovanidis, L.D. Alvarez-Corrales, and L. Decreusefond. Analyzing interference from static cellular cooperation using the nearest neighbour model. *WIOPT-SPASWIN, Mumbai, India.*, 2015.
- [GB13] A. Giovanidis and F. Baccelli. A Stochastic Geometry Framework for Analyzing Pairwise-Cooperative Cellular Networks. *arXiv:1305.6254 [cs.IT]*, May 2013.
- [GKB12] A. Giovanidis, J. Krolkowski, and S. Brueck. A 0-1 program to form minimum cost clusters in the downlink of cooperating base stations. *Proc. of the Wireless Communications and Networking Conference (WCNC), Paris, France*, Apr. 2012.
- [GLM<sup>+</sup>07] A.-T. Gai, D. Lebedev, F. Mathieu, F. de Montgolfier, J. Reynier, and L. Viennot. Acyclic preference systems in P2P networks. *Euro-Par 2007 Parallel Processing, Lecture Notes in Computer Science*, 4641:825–834, 2007.
- [GS62] D. Gale and L.S. Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69, no. 1:9–15, Jan. 1962.
- [HHP06] C. Hoffman, A.E. Holroyd, and Y. Peres. A stable marriage of Poisson and Lebesgue. *The Annals of Probability*, 34, No. 4:1241–1272, 2006.
- [HM96] O. Häggström and R. Meester. Nearest neighbor and hard sphere models in continuum percolation. *Random Structures Algorithms*, 9:295–315, 1996.
- [Irv85] R.W. Irving. An efficient algorithm for the “stable roommates” problem. *Journal of Algorithms*, 6:577–595, 1985.
- [KMN06] I. Kozakova, R. Meester, and S. Nanda. The size of components in continuum nearest-neighbor graphs. *The Annals of Probability*, 34, no.2:528–538, 2006.
- [LMJLH15] N. Lee, D. Morales-Jimenez, A. Lozano, and R.W. Heath. Spectral efficiency of dynamic coordinated beamforming: A stochastic geometry approach. *Wireless Communications, IEEE Transactions on*, 14(1), 2015.
- [LMV<sup>+</sup>06] D. Lebedev, F. Mathieu, L. Viennot, A.-T. Gai, J. Reynier, and F. de Montgolfier. On using Matching theory to understand P2P network design. *INRIA Technical report 00121604, version 2*, 22 Dec. 2006.
- [LP10] G. Last and M.D. Penrose. Percolation and limit theory for the Poisson lilypond model. *arXiv: 1008.0769v1 [math.PR]*, 4 Aug. 2010.
- [NMH13] G. Nigam, P. Minero, and M. Haenggi. Coordinated multipoint in heterogeneous networks: A stochastic geometry approach. *In: Proc. IEEE GLOBECOM Workshop on Emerging Technologies for LTE-Advanced and Beyond 4G (GLOBECOM-B4G’13), Atlanta, GA*, Dec. 2013.
- [NMH14] G. Nigam, P. Minero, and M. Haenggi. Coordinated multipoint joint transmission in heterogeneous networks. *Communications, IEEE Transactions on*, 62(11):4134–4146, 2014.
- [PGH08] A. Papadogiannis, D. Gesbert, and E. Hardouin. A dynamic clustering approach in wireless networks with multi-cell cooperative processing. *IEEE ICC*, 2008.
- [PLH16] J. Park, N. Lee, and R.W. Heath. Cooperative base station coloring for pair-wise multi-cell coordination. *Communications, IEEE Transactions on*, 64(1), 2016.
- [SH14] A.H. Sakr and E. Hossain. Location-aware cross-tier coordinated multipoint transmission in two-tier cellular networks. *Wireless Communications, IEEE Transactions on*, 13(11), 2014.
- [TKBH12] L. Thiele, M. Kurras, K. Börner, and T. Haustein. User-aided sub-clustering for CoMP transmission: Feedback overhead vs. data rate trade-off. *Proc. 46th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA*, Nov. 2012.
- [TSAJ14] R. Tanbourgi, S. Singh, J.G. Andrews, and F.K. Jondral. A tractable model for non-coherent joint-transmission base station cooperation. *Wireless Communications, IEEE Transactions on*, 13(9):4959–4973, 2014.

- [Wei] Eric W. Weisstein. Circle-circle intersection. *From MathWorld—A Wolfram Web Resource.*  
<http://mathworld.wolfram.com/Circle-CircleIntersection.html>.
- [Zon98] C. Zong. The kissing numbers of convex bodies - a brief survey. *Bull. London Math. Soc.*, 30:1–10, 1998.